DRAFT

Kallio Eeva

Training of students' scientific reasoning skills

University of Jyväskylä

1998

ABSTRACT

Kallio Eeva
Training of students' scientific reasoning skills
University of Jyväskylä, 1998. 90 p.
(Jyväskylä Studies in Education, Psychology and Social Research,
ISSN 0075-4625; 139)
ISBN 951-39-0074-6
Yhteenveto: Korkeakouluopiskelijoiden tieteellisen ajattelun kehittäminen.
Diss.

The main empirical aims of the study were to examine the immediate, delayed and transfer effects of training reasoning in a group of university freshmen. The theoretical basis of the dissertation is Piaget's theory of causal and metacausal reasoning. The philosophical root of causal thought in the so-called 'Galilean' tradition of the philosophy of science is also discussed. The subjects were undergraduates (N=101) of the University of Jyväskylä, Finland. There were two experimental groups, who underwent different training procedures. In the first experimental group (EXP-CV) the Ss were taught a formal control-of-variables strategy of thought. In the second experimental group (EXP-M-CV) the Ss were taught CV-strategy and metacognitive representations of CV. The control group did not undergo any specific reasoning training. The training programme lasted for six weeks (12 hours in total). Both the training procedures were based selectively on Adey et al.'s 'Thinking Science' programme. The Ss were pretested with two Science Reasoning Tasks: the Pendulum (SRT-Pe) and Chemicals (SRT-Ch) which are both concerned with the abilities of causal reasoning. The Comparison task (CT) was used to measure M-CV-schema. The effects of training were assessed using immediate posttest and 16 weeks' delayed posttest. The results indicated that the EXP-CV-procedure significantly improved immediate test-performance indicating that basic scientific skills had been learned. The improved performance was sustained in the delayed posttest. A significant effect of training was also found in the transfer-task, indicating immediate and delayed generalization of the cognitive skills to another schema. A significant effect of training was also found in favour of EXP-M-CV in the immediate posttest and the result was found to be stable in the delayed posttest. The results were also analysed by comparing the results from the pretest with those of the delayed posttest (i.e. over the time of 22 weeks). With regard to the delayed effect of the training of EXP-CV, no significant difference was found. However, the effect of transfer across a longer period was found. A delayed training effect was found for the EXP-M-CV procedure, which improved performance in the long run. In summary, the results were interpreted as supporting the validity of both training programmes, but recommendation is given to training causal reasoning with metastrategies to stabilize learning effect in the longer run.

Key words: training study, university students, scientific reasoning, Piaget, formal reasoning, postformal reasoning, traditions of the sciences.

ACKNOWLEDGEMENTS

I would like to express my warmest thanks to Professor Paula Lyytinen for supervising this thesis. Her emotional support, empathy and realism were of value during the process of creating this thesis. I thank Professor Heikki Lyytinen for his great help: his innovativeness was basically the start of my study of this subject. I also wish to thank Professor Jarkko Hautamäki, whose in-depth knowledge of logical reasoning was of considerable importance. I also express my gratitude to my fourth supervisor, Professor Helena Rasku-Puttonen. Dr. Michael Shayer and Dr. Philip Adey from the University of London, Dr. Andreas Demetriou from the University of Thessaloniki and Dr. Michael Commons from the University of Harvard all offered me advice during the course of the study, which proved to be of the greatest value. Dr. Shayer and Dr. Adey have been very helpful in assisting me with the methods of collecting the data. Dr. Demetriou has also clarified some theoretical questions in connection with his post-Piagetian model of reasoning. Also, it has also been a great pleasure to discuss with Dr. Commons many of the questions involved in my work; he has always had innovative interesting ideas. I also thank Dr. Anna-Maija Pirttilä-Backman from the University of Helsinki and Professor Marja Vauras from the University of Turku for critical comments on the manuscript.

I would also to thank all the students who participated in the study, at the Departments of Psychology and Teacher Education in University of Jyväskylä. Without their long-term perseverance this thesis would not have been possible. Mr. Asko Tolvanen from the Department of Psychology of the University of Jyväskylä has given me useful insights into the use of the statistics. Mr. Lauri Viljanto has provided invaluable technical assistance. Mr. Kari Nissinen gave help in statistical questions. Mr. Stephen Lord and Mr. Roger Noël Smith gave assistance in proofreading the English version of the manuscript. I would like to give my warmest gratitude to my closest circle of friends, who have always been warmly supportive. Without their help it would have been impossible to focus on this study. I would like to express my deep respect with following sentence - *'in all these trials, our progress is tested'* as the medieval mystic, Thomas à Kempis, says. And, lastly, I give my warmest regards to all my relatives.

This study was financially supported by the Ministry of Education of Finland, the Faculty of Social Sciences of University of Jyväskylä, the Central Foundation of Student Associations in Finland (OTUS) and the Ellen and Artturi Nyyssönen Foundation; to all of which my thanks are due.

Jyväskylä, 1998
Eeva Kallio

CONTENTS

PART ONE

SOME THEORETICAL CONSIDERATIONS CONCERNING PIAGET'S THEORY OF DEVELOPMENT

1 SCIENTIFIC REASONING AS A THEORETICAL CONSTRUCT

'The higher the level of development, the less likely is its universal attainment' (Blackburn & Papalia, 1992, 152).

1.1 The definition of scientific reasoning in the context of two modes of thought in the history of the sciences

The Aristotelian and Galilean modes of thought in the history of the sciences[1]. Jean Piaget's psychological theory (Piaget, 1967; 1972; 1979; 1983; 1987a; 1987b) of the development of reasoning is so far the most important theory of cognition (Keating, 1980; Vuyk, 1981a; 1981b). Piaget's interest was quite explicitly in the field of the development of scientific reasoning capabilities (Inhelder & Piaget, 1958). The most important reason why Piaget's theory will be described is that it focuses on scientific thinking as *causal* thought (Inhelder & Piaget, ibid.). As will be argued later this is basically the most important developmental form of scientific reasoning supported in the last few centuries in the history of sciences (see, e.g., Dijksterhuis, 1986; Kuhn, 1957; 1970). The final form of the development of reasoning is called formal reasoning (Inhelder & Piaget, 1958). There is also interest in the field of cognitive-developmental psychology to assume a further development of formal thought manifesting itself in adulthood (i.e., a level of so called 'postformal reasoning', see, e.g., Alexander, Druker & Langer, 1990; Kincheloe & Steinberg, 1993; Labouvie-Vief, 1980; Rybash, Hoyer & Roodin, 1986).

According to Georg Henrik von Wright's important philosophical analysis (1971) there seem to be at least two major traditions of science in Western intellectual history: the so-called 'Aristotelian' and 'Galilean' traditions. These historical ways of thought are also relevant in the field of psychology, and the major argument in the following text will be based on von Wright's analysis (ibid.). The claim will be made in this chapter that Piaget's theory (Inhelder & Piaget, 1958) is describing a mode of thought that includes philosophical

1 It should be noted that the word 'sciences' will be used to refer both to the so called 'natural' and 'human' (or humanistic, social and cultural) sciences. The author knows that in the strict sense in the English language 'the sciences' may refer in certain connections only to the natural sciences. However, in this dissertation no such difference is made, and the word is used in a general sense to refer to all sciences, i.e. all academic disciplines which use a system or method based on an agreed scientific consensus. If intentional differentiation is made, it is clearly formulated by stating 'natural sciences' or 'humanistic/cultural sciences'.

assumptions from the Galilean tradition of science. It will be argued that this tradition has had the strongest impact on Western sciences during the last few centuries.

Briefly, von Wright (1971) uses interchangeably the names Aristotelian tradition of science and 'hermeneutics' to refer to what are obviously the same phenomena. It seems that von Wright (1971; 1976) claims that Aristotelism and Galileism have developed in parallel fashion during the history of the sciences[2]. He also states that they cannot be reduced to one united mode of scientific thought, and that they are both needed in every scientific account. Riedel (1976, 7) states that (with an addendum of my own: 'according to von Wright these traditions...) are associated with the opposition between causal and teleological thinking on the one hand, explanation and understanding on the other'. Note that these two traditions of science thus include two different features: *'a different mode of thought'* and *'a different ways of scientific explanation'.* The Aristotelian-hermeneutic mode of explanation is called finalistic or teleological. Briefly stated: this way of explanation, in questions of human behaviour, explains one's action by making the intention of it understandable. There is thus interest in the purposes and aims of human action. The way of reasoning used in this tradition is called *practical syllogism* (von Wright, ibid.).

Galileism has been the most accepted mode of scientific thought and of scientific explanation in the Western world during the last few centuries (von Wright, 1971; see also Heidegger, 1978; Husserl, 1982), an indisputable fact that is still in its heyday (see, e.g., Koch, 1981, with criticism included). There seems to be at least some consensus that Galileism is the tradition of reasoning which has the most important place in the scientific community at present. Philosophical assumptions included in Galileism are (i) an atomistic, variable-centered view of reality; (ii) measurability of the studied phenomena; (iii) causality between the variables; (iv) determinism as a basic feature of reality and (v) predictability based on the lawfulness of reality (von Wright, 1971; some of the above features are mentioned explicitly in the text, others implicitly. See also Kincheloe & Steinberg, 1993; Kramer, 1983; MacDonald & MacDonald, 1995; von Wright, 1976). Galileism is based on the reasoning method borrowed from the natural sciences and the mode of scientific explanation called causal explanation (in contrast to the Aristotelian finalistic-teleological way of explanation) (von Wright, 1971). Philosophically the historically latest form of Galileism is called positivism or logical empiricism (von Wright, ibid.). According to MacDonald and MacDonald (1995) positivism as a philosophical tradition demands the unification of science in a way that all sciences should conform to the model of physics.

It is held in Galileism that reality basically exists as phenomena which

---

2 Care is, however, needed when defining strictly which kind of reasoning is actually called Aristotelian or Galilean. See for example Kurt Lewin's (1935) more than intriguing analysis of these traditions.

may be split atomistically into particles called variables, i.e. reality is seen as separate 'atoms' (Shotter, 1975). Here one makes the assumption that the best way to start to study any system is to 'break it down' in to its basic parts called variables. Variables have to be defined by explicitly formulating the concepts or operations used in the research process. Secondly, measurability was mentioned above as important. It means that any phenomena under study can be described with measurements and calculations, and at the same time it means that scientific language should be mathematical (at least ideally). It also includes the idea of using variables and operations in more or less formal language. Thirdly, the important aim mentioned above is to isolate causal relationships between formulated variables. Causal explanation is our modern way to see reality, which simplistically means that any phenomena under study have their causes which always point to the past. That is, there is a perceived relationship of cause and effect between variables independent of each other. Determinism is assumed in this kind of reasoning: certain variables are determined by the other variables and their relationship is not coincidental. von Wright (1976) defines determinism as something which 'has to (must) be'. Fourthly, the basic interest of any scientific practice is to find generalizable laws in the studied phenomena by finding the above-mentioned relationship between variables (MacDonald & MacDonald, 1995). >Reality as based on a set of laws= is thus seen as a basic feature of the essence of reality; there is no interest in sporadic, irregular or individual features. It is obvious that prediction or at least the prediction of probable manifestation is advanced according to the discovered laws. The control of reality is made possible: laws and causal relationships are found and predictable outcomes may be concluded in advance. According to Kramer (1983) mentioned assumptions manifest themselves as a mechanistic world-view where the *root metaphor of reality is a machine,* meaning that any studied phenomenon is seen metaphorically in an extreme fashion as a technical construct which can be divided into smaller atoms.

So far I have been describing the philosophical presuppositions of the accepted mode of scientific reasoning. The next step will be a psychological description of the reasoning mode explicitly including Galilean assumptions of science. The most important form of scientific reasoning at present, as already mentioned, is causal reasoning (hypothetico-deductive reasoning, e.g., Inhelder & Piaget, 1958). This is the most used and the most important scientific method. The author is aware of the conceptual differences between >hypothetico-deductive method= and >hypothetico-deductive reasoning=, in other words >method= and >reasoning= are not strictly identical phenomena. It is also possible to use the expression the =hypothetico-deductive concept of science=. There is a conceptual difference between =method=, =reasoning= and the >concept of science= but they are here used interchangeably referring to the same phenomena. This is done because of the need for simplification and clarification of a complex matter.

If one changes the terms of the discussion from philosophy to psychology, Jean Piaget's theory of cognitive development (e.g., Piaget, 1967; Inhelder & Piaget,

1958) focuses on the developmental processes of logico-mathematical thought and more strictly on the development of causal reasoning (Piaget, 1967; 1972; 1979; 1983; 1987a; 1987b). There is an obvious connection between causal thought and formal reasoning, the highest reasoning stage Piaget has postulated to exist (Inhelder & Piaget, 1958); thus there is also a link between (i) Galileism as a philosophical tradition, (ii) causal reasoning as a basic form of logical reasoning in this tradition, and (iii) formal reasoning as a psychological construct of it.

1.2 Causal thought and Piaget's psychological theory of formal operations

In a very basic form, 'reasoning' is here defined as a form of thinking which is also called logical thinking. Logical thinking is a way of reasoning which progresses from the given premises to a logical conclusion based solely on the premises (Inhelder & Piaget, 1958). Chains of reasoning are thus used to arrive at conclusions in this kind of thinking. Following the logical rules of reasoning one can make valid conclusions and nothing else: it means that 'with these premises this is the only right conclusion' (Inhelder & Piaget, ibid.). To use a combination of words 'scientific reasoning' gives something more to the definition of logical thinking made above. Arguments were stated in the former chapter in favour of causal reasoning as scientific reasoning.

The development of scientific reasoning as morphological-structural phenomena. According to Flavell and Wohlwill (1969) any developmental psychological theory should necessarily include two viewpoints in describing development: it has to be seen both from the so-called morphological and functional viewpoints. 'Morphology' means that there are fundamental structures in any developing phenomena, or expressed another way, it means that there exists a holistic underlying construct. Structural analysis is needed to demonstrate which kinds of patterns exist behind manifest behaviour. What is also needed is the clarification of the interrelations between the smaller elements of these constructs. What is meant by 'functional' analysis is actually the dynamic which is the basic energy forcing development all the time to go further; it is necessary for producing change in life. Flavell's and Wohlwill's (ibid.) view of development as a simultaneous dynamic and static process is intriguing: there is always interaction between some coherent structure and a dynamic process of change, neither of which are sufficient alone. Essentially the developmental theory of Piaget's fills both conditions: the theory claims that there are structures (i.e. morphological constructs) and processes (i.e. functional dynamics) in cognitive ontogenesis (Piaget, 1967). Cognition is seen as the most developed method of adapting to the environment or, in the other words, it is the latest innovation of *homo sapiens* as a species to adapt to its environment in the biological evolution process. The relationship and interaction between human beings and the environment are thus crucial. So-called *interactionism* is based on this postulate. There is always a relationship between subject and object and fundamentally this interaction is based on action. The basic aim of a human being is to understand reality more and more all the time and in this process continuing development and change are necessary and unavoidable. As a morphological-structural concept *schema* is the basic concept in the Piagetian theory used to describe logical reasoning. It refers to a generalized way of action (Piaget, 1967). Schemata are not empirical in the sense that one can perceive them outwardly, they are psychological in their essence; outer, manifest behaviour is only indirectly referring to them. Depending on the level of development, schemata may be sensori-motor or mental ones. To clarify the concept even more, a schema may be described as a vehicle or way to control any new situation by using the old, already-developed ways of action. Thus, any behaviour is always based on the former experiences. A schema is actually an organized way of action which means that human being has the tendency to stabilize behaviour in using the former ways of activity which has been manifested to be beneficial in an individual's history. Any new action or experience a subject is confronting is *assimilated* into the already-existing schema.

Schemata tend to integrate with each other forming more encompassing totalities called *structure d'ensemble.* It seems thus that there is a tendency to fixation and stabilization forms in cognitive development. The question remains how change and development are possible if there is a basic tendency to structuralization. Piaget uses the concept of *accommodation* in describing the dynamic aspect of modifying the schemata. Assimilation is passive in its essence. It does not demand the alteration of the already-existing ways of action. Schemata can be modified when new experiences are in conflict with formed, existing schemata; i.e. when new experiences are contradictory to previous experiences and cannot be handled with existing capabilities. Change is thus possible by actively modifying the existing schemata to more adaptive forms.

In brief, what is needed in developmental change are contradictory and different experiences in contrast to already-formed ways of action (Piaget, 1967). Piaget (ibid.) calls the relationship between these psychological polarities the *equilibration process* of cognitive development.

Longer periods called developmental stages of cognition are equilibrated states of intellectual development. It is typical to speak about different developmental stages based on the clear qualitative changes between equilibrated states. The sensomotor-, preoperational-, concrete- and formal operational stages of cognitive development are thus defined (Piaget, 1967). Because development is viewed all the time as continuing phenomena, there are always minor substages inside any major stage, and also transitional substages between stages. Although the stages are universal and global, there are always so called 'décalages' when evaluating the performance of individuals. There seem to be differences in human performance from time to time when abilities are measured. According to Vuyk (1981a) and Beilin (1989) the development from one stage to the next one is never linear but more circular in its essence. Thus, care is needed in describing the stages as absolute, strictly-limited constructs. Besides the structural description, functional dynamism has always to be remembered: both progressive and regressive components always exist at the same time in development.

Any higher development is based on the former, developmentally earlier ontogenetic development. So causal reasoning has to be based on the former-developed action schemata and stages developed in childhood and youth. The sensomotoric stage of development lasts for childs the first two years. During this stage, outer manifest actions are the fundamental forms of thought (Piaget, 1967). Psychological links are connected between motoric movements and sensory experiences and the first practical organized ways of action, i.e. schemata, are thus formed. According to Piaget's theory, these actions are a necessary condition for the later development of interiorized logical schematization. Especially important are the outer actions which are *reversible.* (Actually this concept is the core concept of Piaget's theory of development of logical reasoning: later mental logical operations are only interiorized and are just more complex forms of it.) These kind of motoric actions (i.e. the transition of an object from place A to B: A -> B and backwards: B -> A) are thus developmentally the first condition for any further development in reasoning. The second, preoperational stage of thinking lasts from two years of age to the beginning of school-age, i.e. to seven years of age. Inner mental and visual representations of outer reality are the new innovations during this period. But still there is no mental-based logical reversibility; thus thinking cannot be called logical. Operationality as mental reversibility is for the first time possible during the stage of concrete operational reasoning, which lasts from the seventh year onwards to the beginning of puberty, i.e. to eleven or twelve years of age. Inner reversible operations are now for the first time possible, for example classification, and the seriation of objects. The clear lack of reasoning abilities is seen in the fact that reasoning depends quite heavily on empirical facts and there is no ability to handle hypothetical and theoretical possibilities.

A qualitative and equilibrative restructuring of thinking emerge during puberty with the birth of formal operational reasoning (Inhelder & Piaget, 1958; Piaget, 1967). According to Inhelder and Piaget (1958) these following psychological features are seen on the formal operational level of reasoning. (i) One is able to distinguish one from the empirical reality, which means that she or he is able to understand that the empirical fact is the only possibility of the logically multiple possibilities existing conceptually. (ii) At the same time, reasoning becomes systematic. This means that one is able to understand that possibilities always form a logical construct of all theoretically possible combinations. (iii) Realizing the logical possibilities means that one is capable of hypothetical thinking, because potential cases are seen as hypotheses to be studied. (iv) To go even further, experiments are constructed to study the truth-status of the stated hypothesis. (v) Conclusions based on deductive reasoning are made from the empirical facts based on the experiments. Thus, the ability to verify or falsify a hypothesis becomes possible (Keating, 1980; 1990).

In summary, it is obvious that causal reasoning is included when using the concept of formal reasoning. As it is clearly stated in Inhelder's and Piaget's (1958, p. xiii) book: 'in formal operations... most important features are the development to use hypothetical reasoning based on a logic of all possible combinations and to perform controlled experimentation'. Or, as it is expressed in the following way: formal reasoning means 'the appearance of two formerly insignificant behaviour patterns: the formulation of hypothesis, ... and attempts at proof, which consist of determining which of the possibilities in fact do occur' (Inhelder & Piaget, 1958, 58). Shayer and Adey state it briefly (1981, 7): '...a facility to handle abstractions and many-variable problems which is a characteristic of formal operations'. Adey, Shayer and Yates (1989b) describe the difference between formal and concrete reasoning as the difference between the ability to explain some phenomena vs. just describing it.

The control of variables as the basic schema of formal reasoning (CV-schema). As mentioned above, every structural stage of reasoning may be described in smaller components called schemata. Also, formal operational reasoning is basically described as schemata. Inhelder and Piaget (1958; following list based on Shayer & Adey, 1981) have proposed that there exist ten schemata of formal thinking. They are (1) control of variables, (2) exclusion of irrelevant variables, (3) combinatorial thinking, (4) notions of probability, (5) notions of correlation, (6) coordination of frames of reference, (7) multiplicative compensation; (8) equilibrium of physical systems, (9) proportional thinking and (10) physical conservations involving models.

The most important and basic schemata of formal reasoning are the isolation, exclusion and control of variables, which are next described throughly because the training studies which will be described later focus on them. I would prefer to call formal reasoning =variable-centered reasoning= although it may also be called causal reasoning. This is because the isolation of different variables is the basic step in any research and problem-solving. Suppose that one is facing a phenomenon which has been manifested to be problematic and puzzling and has to be solved in one way or another. The first obvious sign of reasoning is to start by separating different variables possibly affecting the situation. Briefly said, it is a question of the ability to perceive and formulate variables relevant to deeper study of the problem. Experimentation with different combinations of variables is a natural step after the separation of variables. An even more important innovation of advanced reasoning is, however, the construction of experiments. Then, one chooses pairs of experiments for further analysis. This implies an understanding of the schema of the *control of variables*. The subject chooses a certain single factor for comparison and keeps all other things equal between compared experiments. One is able to verify the effect of one factor by leaving all the other known factors constant in comparison. Here the operation of comparison is meaningful. The formal operational thinker understands that if she or he has to establish a given relationship between variables, it is important to select certain pairs of variables rather than others. One needs to vary each variable in turn while holding the others constant. It is no coincidence that Inhelder and Piaget (1958, 46) state quite clearly that the schema =all other things being equal' plays an important role in causal reasoning. The schema of combinatorial thinking is also closely linked to the schema of isolation and the control of variables. The control of variables presupposes an understanding of the combinatorial system of variables as a possible hypothesis (i.e. all possible combinations of the given variables).

Besides these steps the exclusion of irrelevant variables will also be manifest. The exclusion of the causally irrelevant factors of a group of given variables is here noted as a developmental innovation; it of course naturally also implies the ability to include the causally relevant factors. This is based on the ability to make a logical deduction of the role of every variable based on the experiments. One is able to demonstrate if some factor is a causal agent, and which other variables have no effect (Inhelder & Piaget, 1958).

1.3 Metacausal thought under consideration: 'postformal reasoning' as a neo-Piagetian concept

> *'The more Piaget's interest turned to higher levels of thought, the more important conscious thinking became'* (Vuyk 1981a, 120).

So far it has been claimed that formal reasoning is the highest developmental stage of scientific thinking. However, as stated in the introduction, there have been scholars who have proposed the existence of a higher developmental stage than formal operations. This proposition implies theoretical changes to Piaget's original theory (Inhelder & Piaget, 1958). Piaget himself had the idea *apriori* that there is no further development after formal reasoning and therefore that there is no postformal stage of development; causal thought would thus be the highest innovation of scientific reasoning (after Vuyk, 1981a). I stress this point, because Piaget has made here a clear philosophical claim. He had an interesting attitude towards philosophy (Piaget, 1971). According to his opinion philosophical claims are not knowledge in the strict sense, and rather that they have only the status of wisdom. The only way to collect true knowledge ('the truth') is by using procedures of scientific verification. Philosophical, speculative reflection can lead to the elaboration of a hypothesis but the criterion of truth may only be thus subjective. According to Piaget (ibid.), there may be several philosophies and several wisdoms of the same facts, but there exists only one truth. As stated above, Piaget has claimed *apriori* that there is no postformal stage of reasoning. Because he has made this claim *apriori,* it is a clear philosophical assumption - he has not made an exhaustive empirical analysis of the existence of a postformal stage of reasoning and he had thus no empirical argumentation behind his claim.

What is not surprising is that there is controversy concerning the nature and existence of postformal abilities (e.g., Keating, 1985; Kramer, 1983). Kohlberg

(1990) has made an important contribution in defining more clearly the problem as a question of whether developmental 'levels' or 'stages' exist in the models of postformal reasoning. He has been sceptical if it is a question of stages in the strict Piagetian sense. Kohlberg=s question of *structure d'ensemble* is still unanswered and it is for this reason that in this dissertation postformal reasoning is carefully called as a level of reasoning instead of as a stage of postformal reasoning.

Briefly, the postformal development of reasoning implies that scientific reasoning will undergo development to higher levels than so far assumed. In itself this hypothesis has been put forward from the beginning of the 1980s using a group of models. Actually, there are at present so many models proposed that it is impossible to focus on them singularly; meta-analysis should necessarily be made even with the risk of losing the individual characteristics of models. For example, Keating (1985) has classified the models in to three groups: (1) those which expand the original Piagetian theory with addendums to it, i.e. no problem is seen in the basic assumptions of the theory but there is only the need for a formulation of new stages; (2) those which explicitly contextualize the theory, i.e. which add to the theory some new psychological elements not originally included in it; and (3) those which confront the theory explicitly, i.e. those models where basic assumptions of the theory in itself are not seen as valid and should be rejected. Research on postformal abilities is widespread and questions of further development in adulthood are crucial indeed. Because of those two facts these models definitely must be taken into account in any study of adult cognition (Alexander, Druker & Langer, 1990; Cavanaugh, 1991; Kincheloe & Steinberg, 1993; Kramer, 1983; Labouvie-Vief, 1980; Mines & Kitchener, 1986). The problem in this field seems to be that more energy has been put into the innovation of new models instead of conceptual analysis and the collection of empirical evidence. It seems obvious that critical discussion is indeed needed in this field.

What follows is a basic classification of postformal models into three =schools= of thought or traditions of thought. It is not attempted here to give an exhaustive picture of the models proposed; only some definitive features are mentioned with examples of the scholars representing each tradition. The classification of models is based on an article by the author and another scholar (Perttula & Kallio, 1996).

Firstly it must be mentioned that some scholars still follow the original Piagetian theory without formulating any new developmental stage. Michael Shayer and his colleagues are a good example of this school (e.g., Adey & Shayer, 1994; Shayer & Adey, 1981). I would describe this position as the 'conservative' tradition of Piagetian theory: no problem is seen in defining causal thinking as the endpoint of development. Scholars of three schools of postformal thought have, however, made changes to the original theory. I would describe these three latter schools as the modern post-Piagetian tradition in contrast to the above-mentioned conservative way of thought. Perttula and Kallio (ibid.) claim that the scholars of the first of these traditions (i) suppose that there are different *philosophical assumptions* in formal and postformal thought (e.g., Kincheloe &

Steinberg, 1993; Kramer, 1983; Linn & Siegel, 1984). These scholars have a tendency to believe that the difference between formal and postformal reasoning is due to differing world-views, and is not about a new equilibrated form of the structure of cognition. If no difference is seen from the structural point of view it means that both in formal and postformal reasoning it is a question of causal reasoning. The difference rather would be in the pre-empirical, philosophical assumptions made about knowledge and reality. According to Kramer (1983), the following assumptions about knowledge are typical to postformal reasoning: (1) the realization of the non-absolute nature of knowledge (relativism); (2) the acceptance of contradiction, and (3) the integration of contradiction into (dialectical) totality (thus postformal thought is called relativistic-dialectical thought; e.g., Kramer, ibid.). By contrast, during the formal stage of reasoning the assumptions of knowledge are absolute, i.e. there is belief that there exists only one solution and truth to every problem. Also the assumption about the core nature of reality is claimed to change from static to dynamic suppositions, i.e. reality is seen during the formal phase as fixed and static, but in the postformal phase there is continual process and change. Thirdly, in formal reasoning the independence of variables and in postformal reasoning the interdependece of variables is assumed (Kramer, ibid.). Briefly, according to Kincheloe ja Steinberg (1993) there actually is a developmental movement from a mechanistic Galilean-Newtonian worldview to a postmodern relativism.

The second school (ii) appeals to go theoretically even further. They integrate the further development of logical reasoning with the *coordinative functions of self* and other fields of personality development (e.g., Edelstein & Noam, 1982; Kramer, 1983; Labouvie-Vief, 1980). This tradition is heterogenous and only brief features will be mentioned in the following. Edelstein and Noam (1982) claim that during adulthood the self begins to coordinate variable-centred thought systems. During youth, the opposite is claimed to be true: logical principles are seen as the highest norms to be followed. The developmental task of adulthood is seen to be the maturation of autonomy, and also in the field of logical reasoning independence of self is supposed to become stronger. This means that formal reasoning becomes contextual and dependent on subjective and societal meanings. The subject himself selects independently the principles which he follows, i.e. the objective logical rules existing are no longer principles to be followed when subjective (emotional and social) factors also take part in reasoning processes (Labouvie-Vief, 1980). Autonomous, self-coordinated thinking is possible when there are the first time notions of many-valued world (i.e. relativism) and an ability to integrate values, thoughts and norms together independently with one=s self coordination (i.e. dialectical thinking).

There are also scholars (iii) who stress the importance of defining postformal thinking as *metacognitive, integrative ability* focusing on systems of variables of causal thought (e.g., Commons, Richards & Kuhn, 1982; Demetriou & Efklides, 1985). Metacognition as mental phenomena is difficult to define precisely and there seems to be variation in the usage of this word (Brown, 1987; Flavell, 1987; Nelson, 1996). On a very simple level metacognition may be defined as the 'reflection on one's own cognition' (or 'cognition about one's own cognition') (Kuhn, Amsel & O'Loughlin, 1988, 3; von Wright, 1992). Kitchener (1983) claims that on the first level, on the level of cognition, individuals for example compute, memorize, read and perceive; and on the level of metacognition one monitors one=s own progress when engaged in first-level tasks.

1.4 Considerations of the development of metacausal thought

The metacognitive-integrative thinking and developmental aspect of this kind of cognition (see, e.g., Hertzog & Dixon, 1994) is the most important factor here to take into consideration as the hypothesis of postformal reasoning. Metacognition focusing on causal reasoning and developmental processes of it has aroused interest during the latest times. Piaget himself has actually made contributions in studying the development of metacognition, see especially his books *'The grasp of consciousness'* (1976) and *'Success and understanding*' (1978). The basic question proposed by him is how one =cognizes= action schemata, or, as Pinard (1986) expresses it, how do subject take charge of their own cognitive functioning (see also Beilin, 1989). Piaget has been interested in how a human being gradually becomes aware of cognitive processes or how the transformation of action schemata to internal concepts is possible during ontogenesis (Piaget, 1976; 1978). The Piagetian concept of 'reflected abstraction' refers to the highest levels of metacognition connected to the highest forms of developing logical reasoning (Brown, 1987). Reflected abstraction is possible when hypothesis testing, evaluation and imaginative abilities of possible worlds and their outcomes have been developed. The last writings of Piaget (1976; 1978) underline the importance of conscious processes in cognitive development, and this is especially so during the stage of scientific formal reasoning. Briefly: 'At the most mature level, which Piaget would prefer restricted to the stage of formal operations, the entire thinking process can be carried out on the mental plane. The learner can consciously invent, test, modify, and generalize theories and discuss these operations with others' (Brown, 1987, 90).

Some neo-Piagetian scholars have also had an interest in this metacognitive aspect of Piaget's theory (e.g., Commons et al., 1982; and more explicitly, Demetriou & Efklides, 1985). Demetriou and his associates have made important empirical studies on the mechanisms of the change and development of metacognition (as a part of their neo-Piagetian model of cognitive development not detail described here[3]; see, e.g., Demetriou, 1990; Demetriou & Efklides, 1989; 1990). What is of most importance here, is that they have studied also the

3 According to Demetriou and Efklides (1985) formal reasoning develops at two levels (strategy and tactics). The level of tactics is structured by different >capacity spheres=. Metacognition is one of these capacity spheres. Each of the spheres has been claimed to develop through sequences. According to Demetrious= terms, the capacity sphere of metacognition focusing on logical reasoning and its development is of interest here.

highest forms of metacognition. I shall restrict my usage of the definition of metacognition and of its development to the earlier articles by Demetriou et al. (Demetriou, 1990; Demetriou & Efklides, 1985), as in my view, their later contributions have not added theoretically new dimensions to the metacognitive model originally proposed[4].

Demetriou and Efklides (1985, 1070) have defined metacognition as an ability *to reflect on, analyse and contrast one's thought processes, to specify them during task-answering, and lastly, to define the similarities and differences between thought processes evoked by them.* Thus Demetriou and Efklides (ibid.) seem to refer both to the knowledge of the cognitive strategies one is using ('reflect, analyse and contrast one's thought processes'; 'specification of subject's process'), and at the same time they refer to thoughts generally focused on tasks ('definition of the similarities and differences between thought processes evoked by them' (i.e. tasks). Also Demetriou and Efklides (ibid.) claim that subjects *'view the whole set of the ... tasks as different systems that could be analysed and reduced to a more general/abstract rule system'* which clearly refer to metacognition focusing on the properties of tasks as chains of variables and as totality (it must be noted that quite similar claims have also been made elsewhere as a hypothesis for postformal development; see, e.g., Commons et al., 1982; Commons, Goodheart & Bresette, 1996; empirical criticism of this model, see Demetriou & Efklides, 1990; Kallio, 1996; Kallio & Helkama, 1991).

4 Demetriou has later been quite innovative in using other words referring to metacognition as part of his theoretical model. For example, the concepts like 'meta-metacognition' (Demetriou & Efklides, 1990) and 'hypercognition' (e.g., Demetriou & Efklides, 1994a) refer to processes which will be here called metacognition for simplicity. However, a critical attitude has to be kept in mind in this field so as not to make it conceptually even more complex, thus only the conventionally accepted concept of 'metacognition' will here be used.

According to empirical data (e.g., Demetriou, 1990; Demetriou & Efklides, 1985) it seems that there are steps of metacognitive, reflective abilities developing in parallel[5] to logico-mathematical thought. The following levels have been discovered to exist: (i) level of no conscious reflection, (ii) level of content-based reflection, (iii) specification of the cognitive operations involved and (iv) analysis and integration of cognitive operations (e.g., Demetriou, 1990). The fourth sublevel focusing on causal thought seems to be empirically demonstrated to be postformal as the authors claim (i.e. it empirically seems to manifest itself after the full formal reasoning; e.g., Demetriou, 1990). Concept of postformal reasoning is used here only for the forms of metacognitive analysis and integration processes developing after (or during the highest forms of) formal reasoning.

What is interesting from the viewpoint of this dissertation is the definition of a fourth developmental level of metacognition. To put it briefly, the ability to analyse logical-operational processes as smaller components (e.g., as variables, values of variables, and interrelationships between them) is a necessary condition for postformal reasoning. But what is most important, the integration (or reduction) of thought operations to a general rule system is crucially the core and the essence of highest forms of metacognition. The latter capacity needs as its precondition more than one formal-operational system (for example, formal tasks) to be compared, contrasted and evaluated with each other. These integrative properties of metacognitive thought are manifested by clarifying the differences and similarities of thought processes evoked and, what most importantly, reducing them to metaclaim (Commons et al., 1982) or general/abstract rule system (Demetriou & Efklides, 1985). Thus, one is able from the metatheoretical perspective (Biggs, 1992) to analyse the lower-level components and make a holistic integration of the totality of thought processes. Demetriou and Efklides (1994a, 6) say that '...metacognitive experiences are only grossly analysable before reasoning itself is well developed. This occurs in college years when the reasoner becomes a virtual logician; a theorist of reasoning'.

Briefly, the highest forms of metacausal thought is here defined to be (i) causal reasoning in its logical essence, and at the same time (ii) it includes exhaustive metacognition focused on thought processes of reasoning, and (iii) metacognition focusing on analysis and the integration of properties of causal thought. Earlier in this thesis the definition of formal reasoning was restricted to the CV-schema; in the case of metacausal reasoning it is question of metacognition focusing on the CV-schema (and it will be called the *M-CV-schema*) (see Figure 1).

5 This is one of the differences between the models of Demetriou and Commons. For example, Commons claims stages exist beyond formal reasoning during which metacognitive processes become manifest (e.g., Commons et al., 1982, p. 1059). Demetriou is interested in the parallel development of metacognition to logical reasoning (e.g., Demetriou, 1990), claiming that it has its own ontogenesis, separated from reasoning but at the same time, in relationship to it.

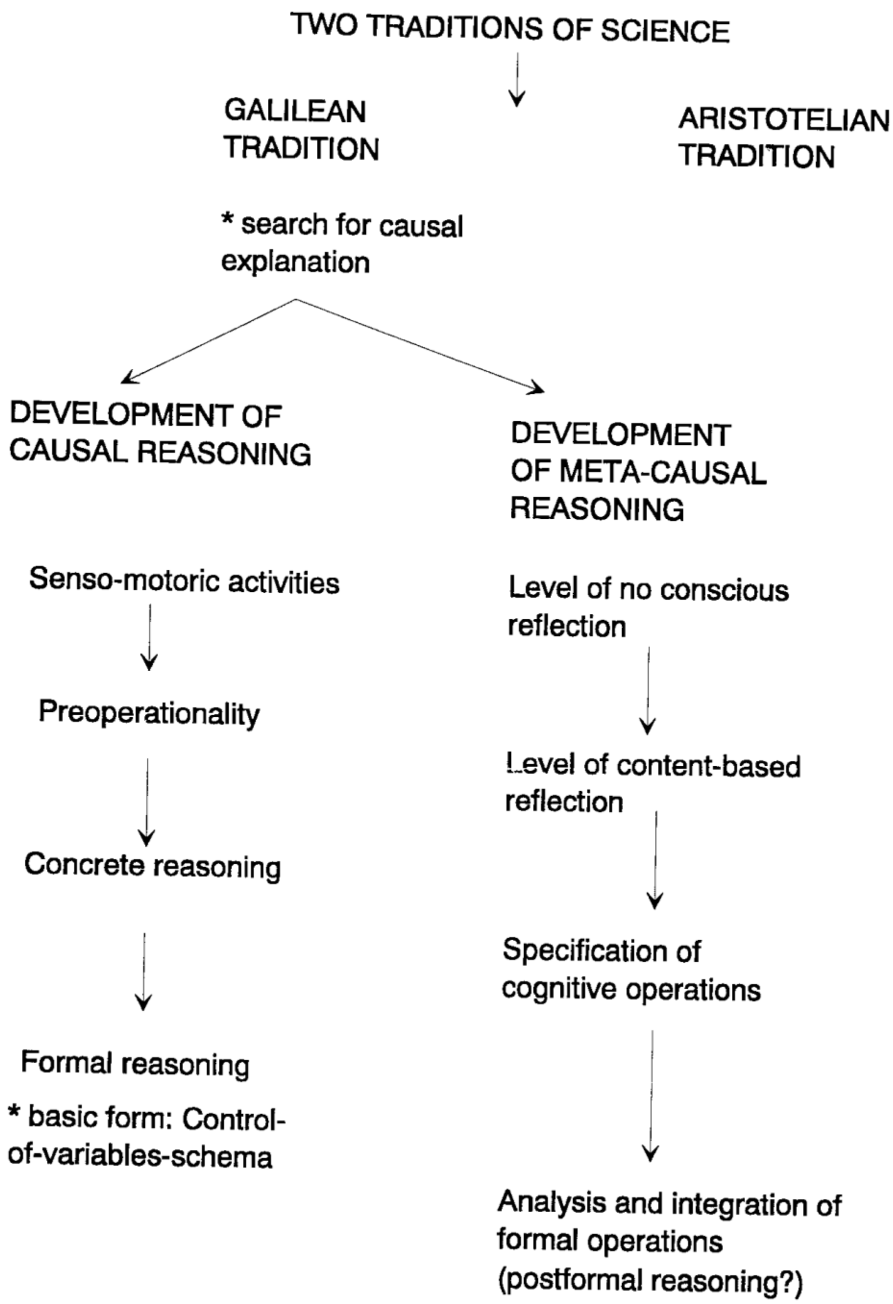


FIGURE 1 The theoretical construct of the study summarized.

PART TWO

EMPIRICAL VERIFICATION OF THE TRAINABILITY OF SCIENTIFIC THOUGHT

2 PREVIOUS TRAINING STUDIES IN THIS FIELD

2.1 The need for training scientific thought among university students

> *'The assumption is often made by college professors that incoming freshmen students think logically'.* (McKinnon & Renner, 1971, 1047).

Formal reasoning and reflection on it can be claimed to be the highest developmental levels of scientific reasoning so far attained, as has been argued in the previous chapters. Well-developed causal thought with metacognitive components is clearly phenomena which are only commanded by a few in our society. Concrete operational reasoning seems to be in general the highest developmental level of logical reasoning among average adults, i.e. it may be called the universal stage of logical reasoning (King, 1986). It is also a fact that not all university students have reached the level of formal reasoning. McKinnon and Renner (1971) claim for instance that only 25% of students are able to reason causally. Also Thomas and Grouws (1984) argue that only 25 to 40% of college freshmen have well-developed abilities of formal operational reasoning. The numbers vary from study to study and a generalized conclusion is difficult to make, but it is clear that formal operations are not in general commanded among students. This is emphasised in a study by Lawson and Snitgen (1982) which found that even the CV-schema is not a schema generally possessed among university students - yet it is the basic schema of formal reasoning. The highest forms of metacausal reasoning are therefore very rare, as is implied in the above facts. Indeed, metacausal thought is only found among the brightest students. According to the previous studies worldwide during the 1980's and 1990's, approximately only 5% of undergraduates are able to think metacausally (Commons et al., 1982; Demetriou & Efklides 1985; Kallio, 1996).

Is it possible to train the highest forms of scientific reasoning, and if it is, by which methods can reasoning be improved (Baber, 1994)? When discussing this question with colleagues and students from various disciplines, there has often been a quite sceptical attitude towards the possibility of raising the level of reasoning intentionally. It often seems that reasoning abilities are supposed to be developed naturally, without interventions. Questions of strategic training to raise the cognitive capabilities of adults are seen as difficult; the belief is held that the capacity to change and develop exists mostly in the earlier years of the life-span. Lately, especially in the field of adult lifespan learning, this assumption has been criticized (Sinnott, 1994). The question of the training of scientific thinking is even more pertinent, as lately there has been discussion about improving the quality of higher education, which implicitly contains beside the other important features the question of improving the scientific reasoning skills of university students. Books concerning the general training of thinking abilities have been published (e.g., Baron & Sternberg, 1987; Coles & Robinson, 1989; Collins & Mangieri, 1992; Nickerson, Perkins & Smith, 1985). However, these have not been concerned with the Piagetian approach to scientific reasoning and thus were not included in this study.

> Some brief words are first needed for conceptual clarification of 'development', 'learning', and 'intervention' (Adey & Shayer, 1994). Actually what has to be first explained is the relationship between development and learning. *Development* refers to phenomena in which is implicitly included the idea of maturation as inevitability. Given certain conditions, e.g. adequate nutrition, absence of disease, certain physical and social environmental factors and genetic make up, human development is inevitable (Adey & Shayer, 1989b). By contrast *learning* is something purposeful which may or may not be happening (i.e. it is not 'inevitable' in the same sense as with development). *Intervention* is here defined as a special sort of learning (Adey & Shayer, ibid.).

Lawson (1985), Ross (1988) and Adey (1989) are in agreement on the three main questions which any intervention programme focusing on causal thought must answer. Firstly, the existence of an immediate training effect should always be demonstrated. Secondly, the question of the generalization of these effects to other fields of psychological competence is of importance. This means that transfer (i.e. generalization of taught matter) should always be measured. Thirdly, the stability effects of the training programmes have to be measured.

2.2 A critique of previous intervention studies

After analysing the summaries (Adey, 1988; Lawson, 1985; Nagy & Griffits, 1982; Ross, 1988), eight internationally published studies were found focusing on training university students' formal reasoning[6]. These were the training studies of McKinnon and Renner (1971), Renner and Lawson (1975), Ross, Hubbell, Ross and Thomason (1976), Renner and Paske (1977), Blake and Nordland (1978), Lawson and Snitgen (1982), Thomas and Grouws (1984) and Wilson (1987) (Table 1). In addition to these studies, Adey and Shayer=s (1994) invaluable book based on the long-term experience of teaching formal reasoning is important although it does not directly concern university students. It is based on the most recently published training programme and is founded on strong empirical evidence and the use of standardized tasks; two criteria which will be later argued to be of crucial importance in this field (Adey, 1988; 1992a; 1992b; Adey & Shayer, 1993; 1994; Adey, Shayer & Yates, 1989a; Shayer, 1992; 1993; Shayer & Adey, 1992a; 1992b; 1993). Metacausal reasoning abilities were defined previously (see the relevant chapters of the theoretical introduction) as belonging to a very high level of

---

6 Internationally unpublished sources, e.g., dissertations, were left out of this summary. Only articles and books were included.

reasoning. However the kind of training studies that defined the highest forms of metacausality as analytic-integrative causal thought were not found. What was needed here were studies in which original Piagetian (Piaget, 1976; 1978) or neo-Piagetian developmental models of metacognition were used. Of the latter Demetriou's model (Demetriou, 1990; 1993; Demetriou & Efklides, 1985) of metacognitive development was selected, but no training studies were found concerning it.

The main results of the intervention studies will be shortly reviewed and critically discussed[7]. The results seem to indicate at face value that practical activities (given various names, e.g., 'IQ-technique[8], 'concrete activities' or even 'Mastermind playing') are effective in training formal reasoning among university students (Lawson & Snitgen, 1982; McKinnon & Renner, 1971; Renner & Lawson, 1975; Renner & Paske, 1977; Thomas & Grouws, 1984). There is, however, an exception here, a study by Blake and Nordland (1978) (a replication of study of the McKinnon and Renner (1971)), where practical laboratory-like exercises were *not* found to be more effective than the absence of no special training. It is also interesting that Ross et al. (1976) found that of the three different teaching strategies of formal reasoning, direct didactic rule-giving was the most effective; indicating that the subjects' own cognitive construction in problem-solving is not as important as the production of the right solutions to problems! Controversy thus exists concerning these results over what kind of teaching is the most effective in training causal reasoning among university students.

A clear inadequacy in almost all of the studies is the lack of testing for the retention effects of training; the only exception is the study of Wilson (1987) which found that improved performance was not stable after a short follow-up period. To study only the immediate effect of training is claimed to be trivial (Adey, 1989); clearly the question of the stability of reasoning improvements is crucial in future studies. Adey (1989) also points out that what is of importance is to measure how the learned skills are generalized to fields other than which have been taught. Transfer effects have been taken into account in two studies (Lawson & Snitgen, 1982; Ross et al., 1976), with contradictory results. Transfer was found in Ross et al.'s study of CV-training to near-schema of formal reasoning (to the Chemicals task) but not to the schema of compensation (to the Balance beam task). The delayed effects of training formal reasoning have only been studied in Wilson's (1987) study, where a significantly improved performance was not sustained after a 11-week retention period. Analysing the results more carefully, it was possible to identify also the following topics.

Characteristics of the Ss under study. It may sound naïve to state that in any

7 Please note that there are many general training studies of formal reasoning (Ross, 1988), but only a few concerning university students. The younger age-groups have been studied a great deal, as is demonstrable in Ross's (ibid.) meta-analysis.

8 The IQ-stategy is defined (e.g., Ross et al., 1976; Lawson & Snitgen, 1982) as a teaching technique where practical experiences in laboratory-like environments are made possible. Equipment has been arranged in laboratory-like situations where the students have been asked to explore and solve problems.

scholarly study the basic characteristics of the Ss under research should be reported. The age and gender of subjects, their area and stage of studies are fundamental facts which should be given. Some scholars (e.g., McKinnon & Renner, 1971; Renner & Paske, 1977) however, haven=t even provided basic data concerning the characteristics of their subjects. Those who have done so (e.g., Ross et al., 1976), have indicated that the Ss have mostly been young adults (19-23 years) and from the socio-humanistic fields of the sciences. In others the Ss have been only females (e.g., Renner & Lawson, 1975), and there are also some studies in which both females and males have taken part (Ross et al., 1976); while some scholars, however, have not provided any information concerning the gender of their subjects (e.g.,Wilson, 1987). On occasions groups of Ss have been preselected on the basis of their operational abilities, for example concrete-operational subjects in Ross et al.'s (1976) study and transitional ones in Wilson's (1987) study. Other scholars have not preselected their subjects using criteria based on operational abilities (see chapter 4.3 for definition of operational levels).

Characteristics of the interventions and design. The basic characteristics of the interventions have also been under reported, the earliest studies of which (e.g., McKinnon & Renner, 1971; Renner & Lawson, 1975) are >good= examples. In many cases it is impossible to understand what the interventions actually consisted of, as only a brief description is given, such as 'using the IQ-strategy in the intervention'. Other studies (e.g., Lawson & Snitgen, 1982; Wilson, 1987) have been more promising in their reporting style, in which, as needed, step-by-step details of the teaching sessions have been reported. A published form of the training programmes is also essential, and information concerning their exact content is crucial if there is to be practical application of them. According to Nagy and Griffits (1982) training studies have borne little fruit in prescribing instructional strategies to accelerate intellectual abilities. If the contents of teaching sessions are not published, there can hardly be expected to be any practical use for them. Possibly not surprisingly, the length of intervention and frequency of the teaching sessions used are also missing in some studies (McKinnon & Renner, 1971; Renner & Lawson, 1975). Some scholars provide information of a superficial kind (intervention 'lasting one semester') (Blake & Nordland, 1978). Wilson (1987), and Thomas and Grouws (1984) are exceptions to this rule, giving specific details of their interventions, which in both cases consisted of a short (four to five weeks) intervention period.

The tasks used in pre- and posttests. The psychometrical characteristics of the tasks haven=t been reported in several studies and this was especially so of those made during the 1970s (McKinnon & Renner, 1971; Renner & Lawson, 1975; Renner & Paske, 1977; Ross et al., 1976). Later studies have been more promising in this respect (Blake & Nordland, 1978; Lawson & Snitgen, 1982; Thomas & Grouws, 1984; Wilson 1987). What is crucial here is the manner in which the reliability data, inner coherence and interscorer agreement of the tasks has been reported. The basic features of the tasks should be taken more seriously in future studies.

TABLE 1 Summary of the training studies of formal reasoning among university students. Listed according to a year of publication.

Author(s), year, (N), characteristics of subjects under study (age, stage and area of study, gender, nature of a sample)	**Focus of the study and nature of the intervention (in addition, length of the intervention and frequency of the teaching sessions)**	**Main results (I=immediate testing, D=delayed testing, T=transfer measured)**	**Tasks used in testings. Reported characteristics of the tasks (reliability, inner coherency, interscorer agree- ment)**
McKinnon & Renner 1971 (N=131) - freshmen - females/males	- inquiry-oriented (IQ) science course vs. CTRL - no specific information given concerning the training content	IQ > CTRL (I) (no D) (no T)	- 'Five Piagetian tasks' - no characteristics of the tasks reported
Renner & Lawson 1975 (N=37) - mean 19-years old - undergraduates - teacher education - mostly females	-IQ vs. CTRL -'working with concrete materials'	No difference between the groups after one year, but mean gain in level IQ > CTRL (D?, no I). (no T)	- 'Six Piagetian-style tasks' with interviews - evaluator comparison: no significant difference between the scorers
Ross, Hubbell, Ross & Thomason 1976 (N=57) - about 22 years - males and females - psychology students - selected sample of concrete operational Ss	- training of CV-schema - three groups differing in teaching strategy (cognitive conflict C; key-concept finding K; didactic rule giving D) vs. CTRL - dealing with concrete materials	D > CTRL (I) (T found to chemicals but not to balance beam) (no D)	- adapted version of the 'Tisher test of operational thinking' - no characteristics of the tasks were reported
Renner & Paske 1977 (N=151?; unclearly expressed) - non-majors in science (physics), teacher education - predominantly females	- concrete instruction (C) and traditional method (T) vs. CTRL - dealing with concrete materials - 'one five-semester-hour course'	C > T (I) (no D) (no T)	- 'Piagetian tasks' with interviews - no characteristics of the tasks were reported

(continues)

TABLE 1 (continued)

Author(s), year, (N), characteristics of subjects under study (age, stage and area of study, gender, nature of sample)	**Focus of the study and nature of the intervention (plus length of the intervention and frequency of the teaching sessions)**	**Main results (I=immediate testing, D=delayed testing, T=transfer measured)**	**Tasks used in testings. Reported characteristics of the tasks (reliability, inner coherency, interscorer agreement)**
Blake & Nordland 1978 (N=97) - freshmen - teacher education of science/maths	- IQS (IQ+ social interaction) vs. CTRL - lasting one semester	No difference between the groups (I) (no D) (no T)	- 'Lawson tasks' - no characteristics of the tasks were reported (see next study: reliability according Lawson 0.78)
Lawson & Snitgen 1982 (N=72) - mean 23 years - freshmen to seniors - a teacher education/ biology course	- IQ (+'specific provisions'), no CTRL; pre/posttested Group 1 and only-posttested Group 2; - lasting one semester - IQ-oriented laboratory experiments, discussion, comparing the similarities & differences between the experiments	IQ 'successful' (I) nonspecific transfer not found (T) (no D)	- 'Lawson test of formal reasoning' - reliability 0.78 in original battery, here 0.68
Thomas & Grouws 1984 (N=39) - freshmen to graduate	- Mastermind playing with social interaction and metacognitive questioning (SI), no interaction (NI) group vs. CTRL - One hour/week for four weeks	SI > CTRL (no T) (D unsuccessful)	- 'Renner battery of formal tasks' - reliability .88 - interscorer agreement or inner coherency not reported
Wilson 1987 (N=52) - mean 19 years - freshmen - a foundation course in science - selected sample of transitional Ss	- EXP vs. CTRL - training in the EXP included: a 5-week programme with six hours a week (e.g., tasks, examples, 'extension problems', interview and practice sheets)	EXP > CTRL (I), but not a stable result (D after 11 weeks from the posttest) (no T)	- Pendulum (Shayer) - task characteristics according to Shayer: rel. 79, int. consist. 83 - characteristics according to Wilson not reported; no interscorer agreement

2.3 The 'Thinking Science' training programme and questions concerning the training of metacausal thought

Michael Shayer and Philip Adey have set up an ambitious project for the training of formal reasoning schemata (Adey, 1988; 1992a; 1992b; Adey & Shayer, 1990; 1993; 1994; Adey, Shayer & Yates, 1989a; Shayer, 1992; 1993; Shayer & Adey, 1992a; 1992b; 1993), under the name of the CASE project (Cognitive Acceleration in Science Education), or '*Thinking Science*' as the manual of the programme is called (Adey, Shayer & Yates, 1989b) (the latter name or *TS* will be used in the following when referring to the training programme). The subjects (N=187) at the start of the intervention were 11-12 years old secondary school students. The intervention lasted for two years between 1984-1986. The frequency rate of the interventions was once every two weeks (about two hours at time), in total about 30-35 intervention sessions over the two years. A pilot study was carried out before starting the major intervention project, the results of which were promising. Eight secondary schools were involved in the major study. All the Ss were pretested before starting the training, midtested after one year, post-tested after two yeas and delayed tested after three years. The testing tasks were the Piagetian-based tasks, originally developed already in the 70's (see Shayer's, 1978 dissertation), including reliability and inner coherence data for several formal reasoning tasks. The main results indicate that after the two-year intervention period the experimental group showed a significantly greater gain on measures of formal operational thinking compared with a matched control group. Secondly, the effects seemed to be stable over three years. Thirdly, an intervention transfer-effect was demonstrable, in this case to other formal schemata and also to the other academic fields, i.e., far-transfer was found (Adey, 1988; 1992a; 1992b; Adey & Shayer, 1993; 1994; Adey, Shayer & Yates, 1989a; Shayer, 1992; 1993; Shayer & Adey, 1992a; 1992b; 1993).

The *TS* -programme (Adey et al., 1989b) is a package of materials needed for the carrying out of the relevant teaching sessions. There are 30 different activities included in the manual focusing on different schemata; they may be used all at once or a combination of a mini-set of activities can be used if the circumstances require (Adey et al., ibid., 4). In the *TS,* most of the material needed for its implementation is included in the package (a teacher's guide with instructions and procedure details; pupils' materials and a technicians= guide for construction equipment). The aim is not to provide a totally new curriculum but one kind of enrichment to be included within a typical educational programme. The teaching style is emphasized as a necessary tool for the successful application of the programme (Adey et al., 1989b). The authors stress that just using the programme by itself will not be successful without concentrating on the pupils thinking, in the form of the use of class discussion (pre-, during and post-experiments), cognitive conflict, metacognition and bridging (Adey et al., 1989b, 4-5).

Adey et al. (1989b) have without doubt constructed a training programme of quality. To outline the structure of teaching sessions step-by-step is a considerable contribution, not to mention the existence of a training programme available for future use. However, some questions still need to be asked. Firstly, the question of applicability of the *TS* with older age-groups and secondly, the question of training metacausal thought defined on an analytic-integrative basis using variable-centred systems (e.g., Biggs, 1992; Demetriou, 1990; 1993; Demetriou & Efklides, 1985).

The question of using the activities with older age-groups is an open one. The CASE programme has been utilised with secondary school pupils, as has already been stated. According to the authors (Adey et al., ibid., 4) the programme may be used from years 7-10 onwards. It is obvious that parts of the programme are obviously too childish to be used with older age-groups; thus, modifications of it have to be carefully developed. Obviously common sense has to be used when implementing it in this case. Therefore, the selection of a mini-set of activities from the *TS* suitable for adults, and especially university students, was seen as a challenging question to study. As has been reported earlier, it is possible to use the Piagetian tasks developed by Shayer et al. with adults (Kallio, 1996; Kallio & Helkama, 1991), but what has been missing is the actual use of the *TS* in practice.

The second question proposed also takes into account recent theoretical discussions concerning the more complex forms of causal thought defined in this thesis as metacausal thought. An outline and definition of how to train metacausal thought step-by-step would be a benefit. Demetriou and Efklides' (1985) definition of metacognition has been used in this thesis as an example of a neo-Piagetian definition of this phenomenon (the following definition is also given in the theoretical introduction to this thesis).

According to Demetriou and Efklides (1985, 1070) metacognition is defined as an ability to reflect on, analyse and contrast one's thought processes, to specify them during task-answering and the understanding of the similarities and differences between the thought processes evoked by tasks. Also Demetriou and Efklides (ibid.) claim that subjects 'view the whole set of the ... tasks as different systems that could be analysed and reduced to a more general/abstract rule system'.

A modified version of the *TS* is thus needed, in which a theoretical discussion of the Piagetian/neo-Piagetian concept of the development of metacognition is taken into account to enhance metacausal thought.

3 AIMS OF THE STUDY

The study addressed several questions. Firstly, the aim of the study was to analyse the (1) trainability of scientific reasoning skills. Two specific questions were studied here: the training of causal thought and secondly the training of metacausal thought. In this study the earlier results were extended by also studying the (2) permanence effects of training causal and metacausal thought and (3) the transfer of the training of causal reasoning.

4 METHOD

4.1 Subjects

101 university students participated in the study between the years 1994-1995. All students were from the University of Jyväskylä, and were from the faculties of Social and Education Sciences with fields of study comprising psychology and teacher training (Table 2). Students' ages ranged from 20 to 34 years ($\bar{x}$=23.35, SD=3.11); thus, they were mostly representative of young adults. The sample was strongly in favour of women (86 vs. 14%). All were undergraduate students who had started their studies between the years 1989-1994. As a measure of their basic intellectual capabilities the shortened version of the Raven S-matrices (scales B, C and D) was used (Raven, 1990). The scores on the task varied from 26 to 36 points (maximum score 36) ($\bar{x}$=33, SD=2.48). The subjects participated on a volunteer basis. They were informed that taking part in the training programme would not entitle them to monetary payment, although they could receive one study credit if they so desired.

TABLE 2 Descriptive characteristics of the subjects under study (N=101).

Variable	%
Gender	
1= female	87
2= male	14
Age in years	
1= 20-23	68
2= 24-34	32
Year of university studies	
1= first year	40

2= 2nd-3rd year	51
3= over 3 years	09
Field of study	
1= psychology	77
2= teacher training	23

4.2 Design of the study

A traditional experimental design was used with two training groups and one control group. The design consisted of a pretest *(PRE),* a period of training and two posttests *(POST1, POST2).* The students were randomly divided into experimental and control groups. The experimental groups will be called *EXP-CV* (N=34, 33 %) in which control-of-variables reasoning was taught, and *EXP-M-CV* (N=38, 37 %) in which control-of-variables reasoning and the metacognition focusing on it were taught. The control group *(CTRL)* which received no specific training (N=29, 30 %).

4.3 Tasks of scientific reasoning

Three measures of scientific reasoning were used: 'Science Reasoning Tasks' *(SRT)* called the Pendulum *(SRT-Pe)* (Shayer, Wylam, Kuchemann & Adey, 1978); the Chemicals *(SRT-Ch)* (ibid.) for measuring causal reasoning; and the Comparison task *(CT)* (Demetriou & Efklides, 1985) for measuring the metacognitive awareness of the mentioned schema. When answering the tasks the same order was always followed, namely first the SRT-Pe, secondly the SRT-Ch and lastly, the CT. These tasks took about three hours to answer. Finnish versions of all the tasks were used.

The Science Reasoning Task Pendulum (SRT-Pe). The version of the task used is constructed by Michael Shayer and his colleagues (Shayer et al., 1978). The task was administered on a group-basis using the video. The experimenter showed a video of four experiments using a pendulum. The experiments differed from each other in the combinations of three variables (i.e. swing, weight and push). Based on the given experiments, the subjects had to conclude which kind of causal effect each variable had on the number of swings the pendulum made in a half minute. The task consisted of answering various questions. The responses to all questions were scored dichotomously, either right or wrong, and one point was given for every right answer. The total sum of right answers was summed and converted to points, which corresponded to a number on a scale expressing the developmental level (scale constructed with Rasch scaling, see Adey & Shayer, 1994). The range of developmental levels was from 2B to 3B*, indicating levels from full concrete operational reasoning to formal generalization (see Table 3)

(Shayer et al., 1978)[9]. See Appendix 1 for the tasksheet; Appendix 2 for scoring rules and Appendix 3 for developmental levels according to the scores. For proper usage of the task and the appropriate instructions see the Science Reasoning Tasks (Shayer et al., 1978). All the task-materials used in this study can also be obtained directly from the author.

The Science Reasoning Task Chemicals (SRT-Ch). The task was based on Shayer's dissertation version (Shayer, 1978). This task was administered by video and on a group-basis. The experimenter performed one experiment with four and another experiment with five chemicals on the video, and the subjects had to answer a number of questions relating to the experiments. The focus of the task was to deduce from the given data which of the variables (colourless liquids/chemicals) does have a causal effect on colour-formation in mixing the liquids. The range of developmental levels was from 2AB to 3B* indicating levels from mid concrete reasoning to formal generalization (Table 3)[10]. The scoring and scaling of the developmental levels was technically similar to the SRT-Pe (see Appendix 4 for tasksheet; Appendix 5 for scoring; Appendix 6 for developmental stages). For proper application of the task and appropriate instructions see Shayer's dissertation (1978). All the task-materials used in this study can also be obtained directly from the author.

The Comparison task (CT) of the SRT-Pe and SRT-Ch. The task was constructed in a similar way to Demetriou and Efklides=s (1985) 'metacognitive' task. The subjects had to compare the SRT-Pe and SRT-Ch with each other. There were 2 open questions to be answered: the subjects had to evaluate the task-properties and analyse their own thought-processes during the task-answering (see Appendix 7 for tasksheet). Demetriou et al. (1985) used a ranking system of four levels of metacognitive development. However, in this study, for more sensitive scoring the subjects were ranked into six levels; two intermediate sublevels (3 and 5) between two major levels were added. These sublevels were constructed on the basis of the exhaustiveness of the answers (see Appendix 8 for the scoring rules of the task).

---

9 Behavioral descriptions of successive levels of competence in the SRT-Pe (Shayer & Adey, 1981): **SRT-Pe** (2B*): Identifies the effect of the salient variable length, but cannot produce a valid reason to justify the deduction; (3A) Can produce a plan for controlling all variables but one in testing for each possible effect (control of variables); (3A3B) Can systematically exclude irrelevant variables in analyzing experiments planned at level 3A, and thus can identify the non-effect of push even if this is counter-intuitive; (3B*) Can systematically resist the impulse to interpret experiments where more than one variable has been changed, and can integrate the two strategies of control and exclusion of variables.

10 Behavioral descriptions of the **SRT-Ch** (Shayer & Adey, 1981) (2B*): Can conceive all the combinations of four objects. Produces a qualitative model of two variables being sufficient for an effect. (3A) Produces an exhaustive set of combinations of 4 objects readily. (3A3B) Can draw inferences from the combinations of four chemicals used what are the necessary and sufficient conditions for an effect and its converse. (3B*) Can produce a proof strategy to justify inferences made at the former level.

TABLE 3 Developmental ranges of the tasks used.

Task	Developmental range	
SRT-Ch	2AB	Mid concrete operational
SRT-Pe	2B	Full concrete operational
	2B*	Concrete generalization
	3A	Early formal operational
	3AB	Full formal operational
	3B*	Formal generalization
CT	1	Level of no reflection
	2	Level of reflection of the content of the tasks
	3	Level of developing general analysis
	4	Level of general analysis
	5	Level of developing specific analysis and integration
	6	Level of specific analysis and integration

Reliability of the tasks. According to Shayer and Adey (1981) the SRT-Pe has shown an internal consistency using KR20 (i.e. alpha coefficient) .83; and the SRT-Ch one of .76. It should be noted that Hautamäki (1984) obtained an identical reliability for the SRT-Pe (.83). Shayer and Adey (ibid.) also report test-retest correlations: with the SRT-Pe .79 and with the SRT-Ch .64. Task-interview correlations of .71 and .65 have been found with the SRT-Pe and the SRT-Ch respectively. In this study the internal consistencies using an alpha coefficient of KR20 were .81 and .53 for the SRT-Pe and SRT-Ch respectively. The test-retest correlation was measured concerning the data in the CTRL group between PRE and POST1 (i.e. among Ss who were not under any treatment). The test-retest correlation with the SRT-Pe was .55 and with the SRT-Ch .61. The item-intercorrelations of the tasks were also used as an indicator of their inner coherence. The intercorrelations of the items varied from -.07 to .85 for the SRT-Pe and for the SRT-Ch from -.14 to .68. The task-intercorrelation between the two formal tasks was .22.

Contrary to the psychometrical strictness of the SRT-tasks, there were difficulties in finding similar evidence for the CT. This task has not been in extensive use and thus the shortage of reliability data is understandable. Demetriou and Efklides (1985) do not report any reliability data for their metacognitive task. In this study, as an indicator of reliability a test-retest correlation was used (in the CTRL between PRE-POST1); the correlation was .42.

Validity of the tasks. The question remains if the tasks measure theoretically the phenomena under study. It is obvious that the SRT-tasks are theoretically directly based on Inhelder and Piaget's (1958) theory concerning formal reasoning. The SRT-Pe focuses on the schema of the isolation and control of variables, again identical to Inhelder and Piaget's task (1958). The only distinctive feature is that characteristics of the tasks are more known with Shayer et al.'s tasks. A similar situation exists with the SRT-Ch; both in it and in Inhelder and Piaget's (1958) version of the task, the schema under study is combinatorial thinking and logical deduction from the given premises. In summary, it is obvious that there is a theoretical validity between the original theoretical construct and the used SRT-tasks. With the CT, the theoretical background is based on the neo-Piagetian model of development of the metacognitive processes. The task used was structurally comparable to the original version of the task used by Demetriou and Efklides (1985). The structure of the CT and the original version of the task were similar although the content was not (i.e. this refers to the fact that the author used the SRT-Pe and the SRT-Ch tasks as evaluation-objects in the CT; Demetriou and Efklides (ibid.) have not used these but other formal tasks of which similar psychometric characteristics are not known).

The interscorer agreement. Using correlations as an indicator of agreement, with the SRT-Pe it was .95, in the SRT-Ch .91 and in the CT .81. The other scorer was an undergraduate student of psychology, with whom the author analysed 20 randomly selected answers from all three tasks. Disagreements in scoring were resolved by discussion between the two scorers.

4.4 Training programmes

The training programmes in both the EXP-CV and EXP-M-CV were based on Adey et al.'s (1989b) *TS* programme. In the EXP-CV the treatment was based on the original version of the training programme. In the EXP-M-CV a modified version of the programme was used. In the EXP-CV the training was designed to induce a schema of control-of-variables and in the EXP-M-CV it was designed also to induce metacognitive awareness of the mentioned schema.

'Activity' refers here to the totality of one teaching session, i.e. to about two hours' time. The *TS* package includes (i) a teacher's guide (i.e. a manual needed for the proper application of the programme), (ii) students' materials (i.e. work-/tasksheets) and, (iii) a technician's guide to construct the equipments needed in the demonstrations (Adey et al., 1989b). As claimed in the theoretical introduction, the basic schema of formal reasoning is the CV-schema[11]. Six teaching activities (sessions) were selected from Adey et al.'s package using criterion focusing on the mentioned schema. The point was to offer only an introductory course to teach basic resources of reasoning. The selected sessions from the mentioned programme were activities 1, 2, 3, 5, 10 and 14 (Adey et al., 1989b) (called in following as 1st-6th teaching sessions)[12]. See Appendix 9 and

11 The criterion for selection was that the CV-schema in itself is demanded as a basic cognitive vehicle in any academic discipline; also on the field of psychology.

12 Note that the activities in the programme itself are focused on natural scientific activities, which means that all the activities are from the fields of physics and chemistry. This reservation has to be kept in the mind in selecting the activities for other fields of studies. Secondly, it has to be

Appendix 10 for structure of every teaching session.

All the teaching sessions followed a similar structure (Adey et al., 1989b) (see Figure 2). Every session had a specific topic; the first part of every session has thus to be *theoretical introduction* to the subject matter: i.e. the teacher's task is, as in any typical teaching session, to give a brief introduction of the basic concepts. For example in the fourth teaching session the relationships between the variables are taught as focusing on the schema of combination of variables. Only the basic concepts have to be defined verbally; formal schemata are never taught directly (thus no direct rule-giving to solve problems is given).

EXP-CV

Activity 4, topic: 'Combinations'.
- **Schema**: Combination of variables

Structure of teaching session:

1. Theoretical introduction of the subject matter by teacher

2. Demonstrations (experimentations)
- practical activities shown on the video taken from the *TS*

3. Working examples (filling tasksheets)
- examples taken from the *TS*
  - >Spinner=
  - >Party=
  - >Switch box=
  - >Germination=
- class discussion of the right answers

4. Bridging taught content to the field of study of students
  - >Four-story problem= by Commons et al. (1982) (logical structure of one of four stories demonstrated)
- class discussion

EXP-M-CV

1.-4. Structure of teaching session as in the EXP-CV

5. Phases of metarepresentation

5.1. The studying of the following article:
- >Four-story problem= by Commons et al. (1982)

5.2. Representation of the structure of the variables in the four stories

5.3. Answering and discussion of the question of the logical similarities and differences between the stories

5.4. Creation of a >metarepresentation= of the stories (short description or meta-statement)

- class discussion

FIGURE 2 Structure of the fourth teaching session in the EXP-CV and EXP-M-CV.

---

noted that originally the programme was developed for the needs of secondary school pupils. It was immediately obvious that all the activities cannot be used in an identical form with students in higher education. Some activities are obviously too childish for adults; changes have to be made before any training undertaken.

After teaching the basic concepts, each session must include *demonstrations (experimentation) of the subject matter*. The intention is here to give food for thought in a manner which will force formal problem-solving during the session. The basic Piagetian intention is that the activities should be practical and this is the case with Adey et al.'s (1989b) training programme. In the version now used of the *TS* the practical activities were shown on the video. Insight with problem-solving is the ideal outcome of this phase of the teaching programme; cognitive conflict is implicitly included in the teaching sessions fostering also cognitive restructuring (Adey et al., 1989b). The teachers' task is to foster experimentation, questioning and the production of cognitive insights. Interaction between the students, and between the students and the teacher, is also crucial. Besides experimentation and social interaction, the teacher should use a *metacognitive questioning technique* at all time during the session. Most important, throughout all the sessions, group-discussions and verbal interaction between the teacher and the students are crucial. The teacher's ability to use discussion technique is paramount in using the *TS*-programme. 'Discussions play a central role in the development of thinking' say the authors (ibid., 4). Adey et al. (1989b) stress this point, indicating that the teaching style of the teacher is important. They conclude that successful using of the *TS*-programme is based on the ability of the teacher to create an atmosphere which stimulates discussion and interaction within the group. Included in this discussion technique is the raising of the students own awareness through questioning. These questions may vary, of course, but in general they are of the type of 'could you please explain... how you arrived at this conclusion?'. Questions should include those such like 'Please explain to the others in your group why you think that?' or 'How did you solve that?' to help focus their attention on their own thought processes. Thus, an important trick for developing thinking skills is for pupils to become conscious and to articulate the sort of thinking they are using to solve problems (Adey et al., 1989b). Indirectly, metacognitive awareness is thus developed also through the teaching style in each training group[13].

Some *working examples* are usually included after the practical demonstrations connected directly either to experimentation or as independent examples. These tasksheet-problems are included in every teaching session; in the fourth teaching session one of the working examples was >Switch box= (Figure 2)[14]. A

---

13 Adey and Shayer (1994, 10) state that 'It is not easy to illustrate this metacognitive element from *TS* activities, because it is more a feature of the teacher's strategy introduced through staff development programmes than of the printed materials'. It must be used as a teacher's technique to provoke self-reflective processes. For analyze these features more carefully, *observations of learning environment* were made step-by-step of every teaching session by two outside observers. The main focus of observations were the quality and quantity of class-discussion and the metacognitive questioning style of the teacher. The observations were made by using Likert-type scale in time intervals of ten minutes. Also questionnaires concerning the experiences of teaching were recorded individually by each student after every teaching session. The data has not been analyzed yet and was thus not included in this thesis.

14 A special box with a lamp, four switches and tester button was constructed with laboratory technician. The tester button lights the lamp when switches 1 and 3 are both >on=, switch 2 has no effect; switch 4 cancels the effect of 1 and 3 (Adey et al., 1989b, Teacher=s guide, 26). The apparatus was demonstrated on the video with several combinations of variables.

sample of questions involving Switch box had to be solved by the students. These puzzles focus on the further interiorization of formal reasoning abilities, in this case, schemata of combination and control of variables. Again, the students own capacities are needed in this session. Metacognitive questioning, discussion and the giving of right answers must of course be included here.

In the both experimental groups the *bridging* was simply an analysis of one article as a system of variables. Examples of ways of bridging are not given in the articles of the CASE, a fact that is special, leaving the way open for teachers to make these kinds of linkages (Adey, 1988; 1992a; 1992b; Adey & Shayer, 1993; 1994; Adey, Shayer & Yates, 1989a; Shayer, 1992; 1993; Shayer & Adey, 1992a; 1992b; 1993). Enhancing the students' understanding of the connections of taught material to other fields of life is an essential feature of the teaching style. It is a way of generalizating what is taught to other fields. It is to do with binding, connecting or making linkages of the teaching matter to other areas of life. Adey et al. (1989b, 5) define bridging in the following way: 'Bridging means simply the linking of the ideas developed in the *TS* activities to parallel experiences in the rest of the science curriculum, and beyond to everyday experiences. Without this sort of mental bridging back and forth, to and from, the *TS* activities there are a danger that they will remain as peculiar or special, and then any effects that may be achieved will remain specifically linked to certain types of activity and not generalized to science, or beyond the science.' As has been already stated, in this study the bridging was made by demonstrating to students how variable-like way of reasoning is obvious in psychological/educational scientific articles.

After the bridging phase of the teaching cycle there was a major difference between the way of teaching in the EXP-CV and the EXP-M-CV. I will call this last phase in the EXP-M-CV as the phase of *metarepresentations* to train specifically metacausal thought in this group. In the EXP-M-CV, a more complex variable-structure was created as metarepresentation than in the EXP-CV. Firstly, the same article as in the bridging was also analysed into components. Secondly, other articles, examples or figures taken from the other studies were also given to students. Thus two or more scientific examples had to be compared with each other in the EXP-M-CV (see Appendix 10 for list of articles used in bridging and metarepresentations). For the analysis of the structure of variables, values of variables and known relationships between them, a separate sheet was given to the students. They had to give also a written formulation of the similarities and differences between the studies. Examples of metarepresentations and used tasksheets for these purposes can be found in the Appendix 11. In this specific case of the teaching session four, an article by Commons et al. (1982) was used as a vehicle of metarepresentation. Four-story-problem (ibid.) was analysed to components: in the EXP-CV one of the stories was analysed, in the EXP-M-CV all four stories were analysed, i.e. demonstrated as chains of variables, and verbal synthesis was created of the similarities and differences.

Teaching in the CTRL group was unrelated to the treatment situation in the experimental groups but relevant to the students' general science course. The contents of the lecture in the CTRL were focusing on the philosophical assumptions of psychology. There were five topics taught and discussed during the lectures: (1) mechanistic-materialistic assumptions of human nature, (2) biological models of man, (3) models including tendencies to underline existentialistic freewill as basic feature of human nature, (4) models pressing the importance of human beings= social responsibility and (5) models focusing on the importance of self-regulative cognitive reflection as the basic nature of a human being.

The teacher was the same (the writer of this thesis) in every teaching session in each group. The students were allowed to be absent from sessions only once; it was obligatory to attend all the other sessions. The sessions were one week apart at the same time every week. As with the testing methods, all the materials included in both training programmes were in Finnish and are available from the author. The main characteristics of the study are summarized in the Figure 3.

PRE: PRETESTING (N=101)

(1) Measuring of scientific reasoning skills concentrating on the schema of the control-of variables (CV) and metacognitive awareness of the mentioned schema (M-CV)

(2) Using the following tasks focusing on the mentioned abilities:
- CV: Science Reasoning Tasks Pendulum and Chemicals
- M-CV: the Comparison task, where students were asked to evaluate the Pendulum and Chemicals tasks with each other

INTERVENTION

EXPERIMENTAL GROUP, EXP-CV (N=34)	***EXPERIMENTAL GROUP, EXP-M-CV*** (N=38)	***CONTROL GROUP, CTRL*** (N=29)
- training CV-schema - lasting 12 hours (6 teaching sessions) - each teaching session followed a similar teaching cycle - selected activities were taught from ready-made training programme - main methods of teaching: lecturing, video-demonstrations of activities, problem-solving of tasksheets, connecting taught content to scientific studies - discussion technique used throughout the sessions	- training M-CV-schema - lasting 12 hours (6 teaching sessions) - each teaching session followed a similar teaching cycle - selected activities as in EXP-CV, with addendums focusing on training M-CV - main methods of teaching: lecturing, video-demonstrations of activities, problem-solving of tasksheets, connecting taught content to scientific studies and evaluative-integrative analysis of scientific studies - discussion technique used throughout the sessions	- course focusing on the philosophical assumptions of human nature implicitly included in the field of psychology - lasting 12 hours (6 teaching sessions) - main methods of teaching: lecturing and discussions

POST 1: IMMEDIATE POSTTESTING (N=101)

(1) Similar testing procedure as in pretesting.

POST 2: DELAYED POSTTESTING (after 4 months) (N=99)

(1) Similar testing procedure as in pretesting and in immediate posttesting.

FIGURE 3 Main characteristics of the study.

4.5 Data reduction

The data collected was reduced to 47 variables. Briefly, the independent variables in the study are the two teaching strategies (CV and M-CV), and the dependent variable is the level of intellectual development. The statistical analysis of the data included firstly the use of descriptive data and gain scores, and secondly the Anova procedures of repeated measures, using the SPSS-X programme. Gain scores were based on the group means of results in different tasks. Selected pairwise comparisons of the groups were made (Table 4).

TABLE 4 Statistical pairwise comparisons made in the study.

The question of study	Testings contrasted/ the task	Pairwise comparison of groups
1a. trainability of CV *1b.* trainability of M-CV	-PRE-POST1/SRT-Pe -PRE-POST1/CT	-EXP-CV vs. CTRL -EXP-M-CV vs. CTRL; EXP-M-CV vs. EXP- CV
2a. stability of effect of the training CV *2b.* stability of the training M-CV	-POST1-POST2/SRT-Pe -POST1-POST2/CT	-EXP-CV vs. CTRL -EXP-M-CV vs. CTRL; EXP-M-CV vs. EXP- CV
3a. delayed effect of training CV *3b.* delayed effect of training M-CV	-PRE-POST2/SRT-Pe -PRE-POST2/CT	-EXP-CV vs. CTRL -EXP-M-CV vs. CTRL; EXP-M-CV vs. EXP- CV
4. transfer effects of training CV	-PRE-POST1/SRT-Ch -POST1-POST2/SRT-Ch -PRE-POST2/SRT-Ch	-EXP-CV vs. CTRL

Effect sizes (ES) of the training were also calculated. According to Adey and Shayer (1994, 39) '...the effect size is the difference between the mean post-test scores of experimental and control groups given in units of the standard deviation of the control group'. The mathematical formula for calculating ES follows Adey's and Shayers= (ibid.) formulation. Thus, for example, calculating ES in analysing the immediate effects of training (the question of study 1a in Table 4) the formula used was: $(\bar{x}_{EXP\text{-}CVpost1} - \bar{x}_{CTRLpost1}) - (\bar{x}_{EXP\text{-}CVpre} - \bar{x}_{CTRLpre}) / SD_{CTRLpost1})$ where $\bar{x}_{EXP\text{-}CV}$ is the mean of EXP-CV in the first post-test, $\bar{x}_{CTRL}$ is the mean of CTRL in the first post-test; $\bar{x}_{EXP\text{-}CVpre}$ is the mean of EXP-CV in pre-test; $\bar{x}_{CTRLpre}$ is the mean of CTRL in pre-test; SD_{CTRL} is the standard deviation of CTRL in the first post-test). The formula varies, of course, depending on the subject of study (see comparisons in Table 4).

5 RESULTS

5.1 The trainability of two modes of scientific reasoning

5.1.1 Immediate effects of training the CV-schema

The substage frequencies in the SRT-Pe in all testing sessions are shown in Table 5. Distributions in EXP-CV in pretest indicated that three students (9 %) were classified as concrete operational (2B). 15 % of subjects showed transitional abilities (2B*) and 47 % of the subjects were classified as early formal operational (3A). 2/3 (71 %) of the subjects demonstrated either transitional or early formal abilities at the beginning of the intervention. The two highest sublevels (3B, 3B*) are ranked as the most mature forms of formal reasoning; 30 % were ranked at these levels. In conclusion, 11 subjects had a ceiling effect (i.e. representing 3B*) in the PRE. They were randomly distributed across all the groups. These Ss were included in all the statistical analyses.

In Table 6 the means and standard deviations in pre- and posttests are given for all the groups. To test the immediate effects of the training the data concerning EXP-CV and CTRL was compared. The gain of scores between testing in the EXP-CV was 1.02. In the CTRL the gain was 0.24. Statistically significant interaction between group and testing was found: the treatment group EXP-CV was significantly better in its results than the CTRL in immediate post-test ($F(1,62)=15.57$, $p<.001$). The ES was 1.39. Graphical representation of the development of causal reasoning over a period of 22 weeks is demonstrated in Figure 4.

TABLE 5 Number of students at each stage on pre- and posttests. The Science Reasoning Task, the Pendulum.

LEVEL	EXP-CV (N=34)			EXP-M-CV (N=38)			CTRL (N=29)		
	PRE	POST1	POST2	PRE	POST1	POST2	PRE	POST1	POST2
2B	3	0	0	2	0	1	1	0	0
2B*	5	0	1	9	1	0	8	1	1
3A	16	10	6	19	8	10	11	20	13
3AB	6	9	11	4	14	11	8	8	8
3B*	4	15	16	4	14	15	3	1	6

Developmental range of the Science Reasoning Task, the Pendulum:
- 2B Full concrete operational
- 2B* Concrete generalization
- 3A Early formal operational
- 3AB Full formal operational
- 3B* Formal generalization

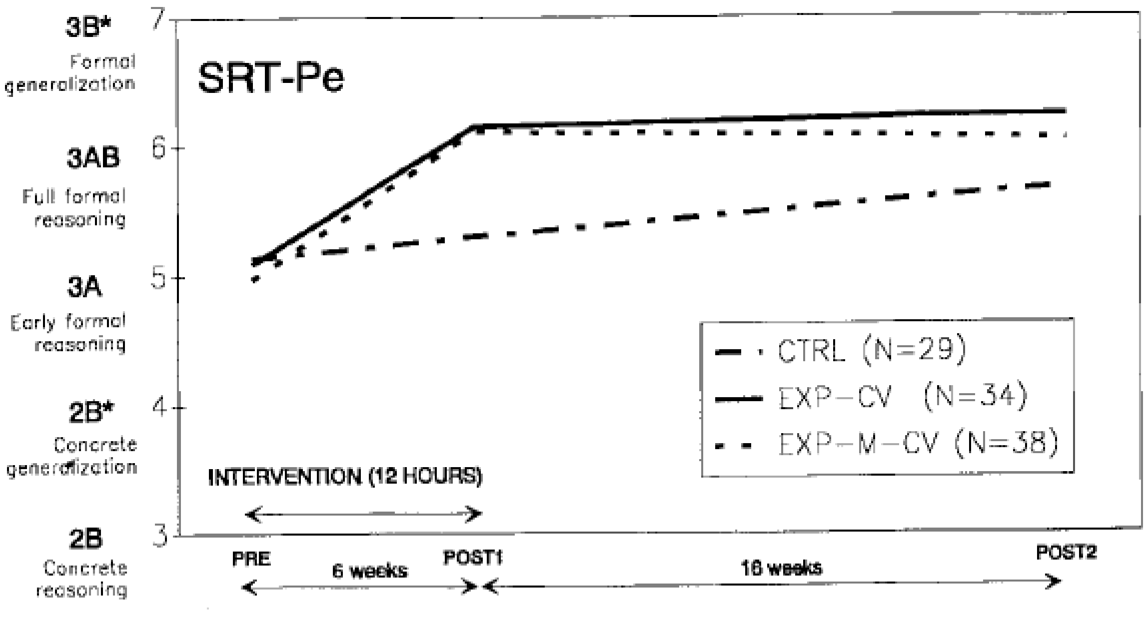


FIGURE 4 Development of causal thought in the EXP-CV and CTRL over a period of 6 and 16 weeks. The Science Reasoning Task, the Pendulum.

TABLE 6 Group means and standard deviations for each task on pre- and posttests.

		PRE		POST1		POST2	
GROUP	TASK	$\bar{x}$	SD	$\bar{x}$	SD	$\bar{x}$	SD
EXP-CV (N=34)	SRT-Pe	5.08	1.08	6.10	0.86	6.20	0.85
	SRT-Ch	5.20	0.72	5.80	0.50	5.60	0.68
	CT	3.50	0.96	3.60	1.12	3.60	0.89
EXP-M-CV (N=38)	SRT-Pe	4.97	1.00	6.11	0.84	6.05	0.97
	SRT-Ch	5.11	0.83	5.89	0.66	5.68	0.67
	CT	3.20	0.97	3.60	0.85	3.80	0.81
CTRL (N=29)	SRT-Pe	5.06	1.02	5.30	0.60	5.60	0.86
	SRT-Ch	5.20	0.80	5.10	0.78	5.10	0.59
	CT	3.29	1.01	2.81	1.01	2.96	0.84

5.1.2 Immediate effects of training the M-CV schema

The distributions as substage frequencies in CT are shown in Table 7. Graphical representation of the metacausal development over 22 weeks is demonstrated in Figure 5. The pretest distributions of developmental levels in EXP-M-CV indicated that 9 % of subjects were classified in the two lowest levels of metacausal development (substages 1 and 2). The intermediate levels 3 and 4 between the lowest and the highest levels represent a general-analysis level of metacausality; 79 % of subjects showed this kind of ability. The two highest sublevels (5, 6) are ranked as the highest levels of metacausal awareness. 12 % (4 students) were able to demonstrate this kind of cognition. One student demonstrated fully developed M-CV-thought in PRE, and she was included in all the analyses.

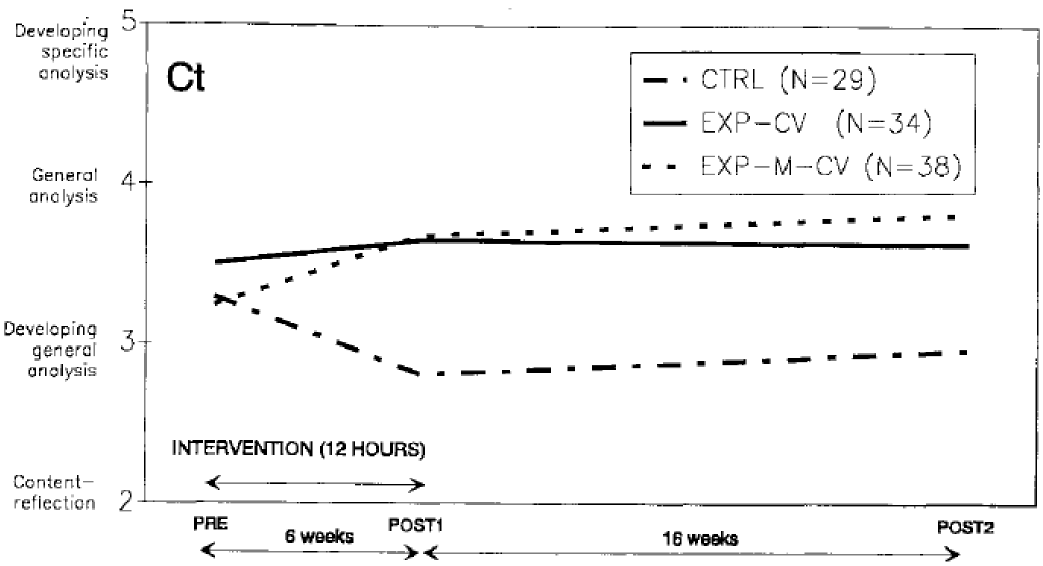
The gain in the EXP-M-CV between testings was 0.4. In the the CTRL the gain was -0.48 (Table 6). Concerning the results of the EXP-M-CV and the CTRL the experimental group was significantly better ($F(1,66)=9.86$, $p <.01$). The ES was .75. Secondly EXP-M-CV and EXP-CV were compared to test for the additional effect of training. The gain in the EXP-CV was 0.1. No statistically significant interaction between group and testing session was found ($F(1,69)=1.22$, p ns); thus, the EXP-M-CV was not significantly better in its results compared to the EXP-CV as control. The ES was .34.

TABLE 7 Numbers of subjects at each stage on pre- and posttests. The Comparison Task.

LEVEL	EXP-CV (N=34)			EXP-M-CV (N=38)			CTRL (N=29)		
	PRE	POST 1	POST 2	PRE	POST 1	POST 2	PRE	POST 1	POST 2
1	1	1	0	2	0	0	3	3	1
2	2	3	2	6	4	1	1	8	6
3	15	11	15	13	9	12	13	14	15
4	12	14	12	15	19	18	12	4	5
5	3	2	4	2	5	5	2	2	1
6	1	3	1	0	0	1	0	0	0

Developmental range of the Comparison Task:
1 Level of no reflection
2 Level of reflection of the content of the tasks
3 Level of developing general analysis
4 Level of general analysis
5 Level of developing specific analysis and integration
6 Level of specific analysis and integration

FIGURE 5 Development of metacausal thought in all the groups over a period of 6 and 16 weeks. The Comparison Task.



5.2 Stability of the effects of training the CV- and M-CV schemata

The gain between POST2 and POST1 in the EXP-CV was 0.01 and in the CTRL 0.57. Stability of performance between posttests was clear in both groups (see Figure 4). A statistically significant interaction between group and testing session was not found ($F(1,60)=3.18$, p ns), thus the developmental change in the groups did not differ from each other. The gain in the EXP-M-CV was 0.20 and in the CTRL 0.15. No significant interaction between group and testing session was found ($F(1,63)=.06$, p ns) indicating stability in the results of both groups which is also visible in Figure 5. Also, EXP-M-CV and EXP-CV were compared to test for the additional effect of training. In EXP-CV the gain was zero. No statistically significant interaction between group and testing session was found ($F(1,69)=1.22$, p ns).

5.3 Delayed effects of training the CV- and M-CV schemata

The delayed effect of training between the PRE and POST2 was also considered. No statistically significant interaction between group x testing session was found in the SRT-Pe ($F(1,60)=3.60$, p ns). A significant interaction effect between group x testing session was found in favour of EXP-M-CV compared to CTRL ($F(1,63)=10.63$, $p < .01$). In comparing EXP-M-CV and EXP-CV a mild linear trend of improvement in the results of EXP-M-CV is clear. A statistically weak significant interaction between group and testing session was found ($F(1,69)=4.37$, $p=.05$), indicating the better learning results of EXP-M-CV over a longer period.

5.4 Transfer effects of training the CV-schema

Distributions of levels in EXP-CV in pretest indicated that 56 % of the subjects demonstrated either transitional or early formal abilities (Table 8). Of the two highest sublevels (3B, 3B*) only 3B was represented in the studied group: 44 % were ranked at this level. There was no ceiling effect in the SRT-Ch.

To test the transfer effects of the training of control-of-variables reasoning the data concerning EXP-CV and CTRL were compared. The gain of scores was 0.6 in the EXP-CV and in the CTRL -0.1. A statistically significant interaction between group and testing session was found ($F(1,62)=11.99$, $p<.001$) indicating that treatment group was able to transfer effect. The ES was .81. The effect of transfer was still visible in the POST2 indicating stability of a result between posttestings (see Figure 6) ($F(1,60=1.60$, p ns). Concerning delayed effects of training, a mild interaction effect in the SRT-Ch in favour of EXP-CV was found ($F(1,60)=4.60$, $p < .05$), indicating the effect of transfer across a longer period.

TABLE 8 Numbers of students at each developmental stage according to the pre- and post-tests. The Science Reasoning Task, Chemicals.

LEVEL	EXP-CV (N=34)			EXP-M-CV (N=38)			CTRL (N=29)		
	PRE	POST1	POST2	PRE	POST1	POST2	PRE	POST1	POST2
2B	0	0	0	0	0	0	1	0	0
2B*	5	0	2	11	1	1	4	7	3
3A	14	7	9	12	7	13	13	12	18
3AB	15	25	21	15	24	20	13	11	7
3B*	0	2	2	0	5	3	0	0	0

Developmental range of the Science Reasoning Task, the Chemicals:
2B Full concrete operational
2B* Concrete generalization
3A Early formal operational
3AB Full formal operational
3B* Formal generalization

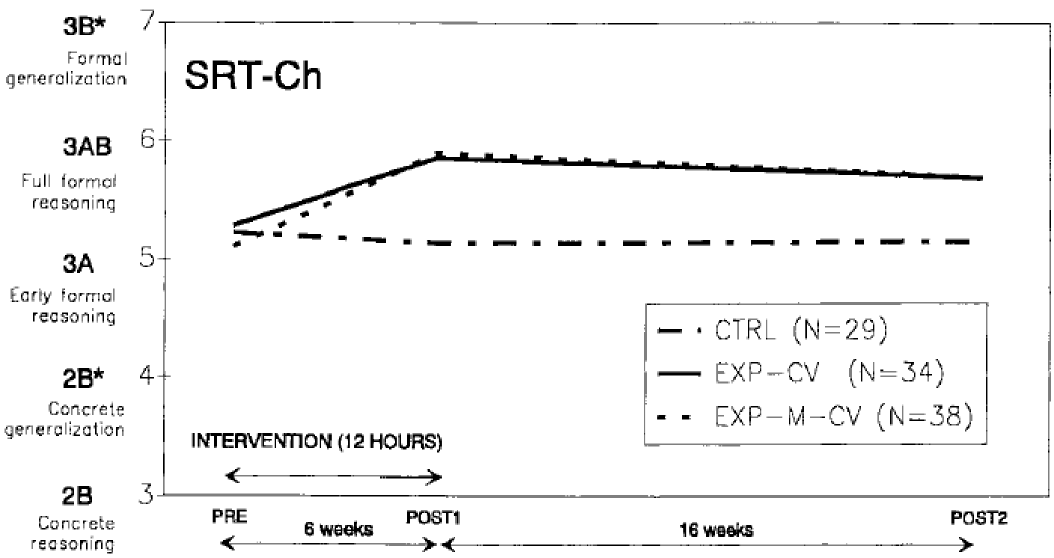


FIGURE 6 Generalization of training CV to another schema of causal reasoning. Cognitive development over 6 and 16 weeks. The Science Reasoning Task, the Chemicals.

6 DISCUSSION

> *'Piaget, in his most recent work, notably in The Grasp of Consciousness (1976) and Success and Understanding (1978) also denotes reflected abstractions, 'Conscious products of reflexive abstraction', as the keystone of formal operations'* (Brown, 1987, 71-72).

6.1 Back to the roots: Galilean and Aristotelian traditions of scientific thought

This thesis was based on Piaget's theory of the development of causal thought, and on a neo-Piagetian model of metacognitive representations of causal thought. It was claimed in the introduction that the theoretical construct of formal reasoning has its philosophical roots in the Galilean tradition of the philosophy of science. So far, philosophical-conceptual analysis like Kramer's (1983), Woods' (1983), and Kincheloe and Steinberg's (1993) has explicated the philosophical assumptions of Piaget's theory. Briefly, it was argued in the introduction that a Galileian mode of thought is the tradition of science most accepted in the scientific community today and that the concept of formal reasoning (Inhelder & Piaget, 1958) is an example of a psychological construct that fundamentally has its theoretical roots in this tradition.

From the viewpoint of the history of science (Kuhn, 1957; 1970), the claim I have stated is actually a paradigmatic one in the Kuhnian sense. Evaluating the change in the history of the sciences from the Aristotelian to the Galilean mode of thought implicitly involves the discussion of the growth, progress and development of the sciences as different paradigms and their interrelationships. From the historical point of view, Galileism was born at around the time when there was change from the medieval world-view to the early modern era, i.e. the period of time from the Renaissance to the Scientific Revolution approximately from 1400 to 1800 AD. (Dijksterhuis, 1986). In the medieval and Renaissance periods, psychology was philosophical in its nature (see, e.g., Kemp, 1990); the change to scientific psychology is a late phenomenon in man=s history. The Scientific Revolution led to the separation of the sciences from theology and philosophy; in psychology the birth of scientific psychology led to a search for a Galilean-causal explanation of psychological phenomena. It has led to a situation where the methods used in the natural sciences have been taken as the necessary foundation for all valid scientific knowledge including in the field of psychology (Feldham, 1995). Feldham also argues that the

Galilean way of scientific reasoning is usually taken in modern Western scientific culture as the paradigm of 'true' scientific accounts but that the Aristotelian way of thought has not been well developed. The same claims are made for the philosophical background of psychology (MacDonald & MacDonald, 1995); see also Shotter (1975) and Winch (1976) for an analysis of the hidden philosophical assumptions about human nature in the Galilean way of scientific thought; Giorgi (1988) for an attempt to construct an 'alternative' phenomenological-hermeneutical method of research in psychology, and Koch (1981) for a general criticism of the situation existing in the field of psychology. See also more recent discussion, springing from Willis Overton's (1994a; 1994b) important articles concerning the paradigm shift in the sciences, and continued with other articles on the same topic (Beilin, 1994; Chandler & Carpendale, 1994; Santostefano, 1994; Scholnick, 1994; Shusterman, 1994). Also excellent is the latest theoretical analysis of presuppositions in experimental psychology by Saariluoma (1997), written from the viewpoint of the cognitive sciences. See also Piaget=s (1983) contribution to this subject.

From the theoretical point of view, the question still remains as to the nature of the interrelationship between the two major traditions of science. The latest discussions studying >children's conceptions of mind= have had connections to this debate. There are proponents who argue that the basic nature of the human mind and human action is essentially intentional (Astington, 1991; 1993; Astington & Olson, 1995; Bruner, 1995; Feldham, 1995; Leadbeater & Raver, 1995). In particular Janet Astington (1991; 1993) has argued in favour of taking into consideration the not-so-favoured tradition of intentionalistic-finalistic thought in psychological research; she has claimed that meaning-making ability of the human's mind is something distinct from naturalistic-causal phenomena (see also Heidegger, 1978; Husserl, 1982). Astington's claims have aroused discussion in which questions put forward originally by Georg Henrik von Wright (1971; 1976) concerning Galileian and Aristotelian modes of thought are being reconsidered in the field of psychology. Briefly, there remains the challenge, as Astington and Olson (1995) state, of how to integrate these two approaches in psychological research.

Analysing, evaluating and integrating the two traditions of science will be essential in the future. Briefly, what is here demanded, is the beginning of a paradigmatic discussion between these two major traditions of science in the field of psychology - not the easiest of tasks but one worth attempting. Integration of different schools of thought in the philosophy of science may be called postformal thinking - essentially it is a matter of the analysis and synthesis of different philosophical modes of thought (i.e. Aristotelian and Galileian mode of thought). Questions of postformal, metacausal thought and metacognitive processes in general might be taken into consideration more carefully in the further discussion of the development of scientific thought. To go even further, to speak about integration of mentioned features of cognition with self-regulative functions (Labouvie-Vief, 1980) and differing world-views (Kramer, 1983) as forms of adult cognition will certainly be discussed in the future also in connection to scientific thought. Also, question of specialized professionalism (i.e. question of experts) in connection to postformal abilities has also to be taken more seriously in further discussion on this field (see e.g., Tynjälä, Nuutinen, Eteläpelto, Kirjonen & Remes, 1997; Eteläpelto & Light, in press).

6.2 Empirical substantiation of training causal and metacausal thought

The tasks. The subjects under study were pre- and posttested with two Science Reasoning Tasks: the Pendulum (SRT-Pe) and the Chemicals (SRT-Ch) which involved the capacity for causal reasoning. The SRT-Pe was used as the criterion task to measure CV-schema, and the SRT-Ch was used as a measure of the transfer of training to another formal schema. The Comparison task (CT) was used to measure M-CV-schema measuring metacausal development. The tasks are promising equipments for obtaining estimates of the cognitive development of university students. The tasks measuring causality (SRT-Pe and SRT-Ch) were found to be quite sound as regards reliability, interscorer agreement and inner coherence[15]. With the CT, reliability and interscorer agreement were found to be good. There seems to be a clear need to develop the CT-task even further to improve the inner coherence of the task. Similar strict standardization to that employed with the SRT-tasks (Shayer & Adey, 1981) is needed here; thus the task should be identically semi-structured, with more questions and items reliably measuring the phenomena under study. Clearly this is a fruitful direction to follow in any further research into metacausal development.

15 It must be noted that in the both tasks some negative item-intercorrelations were found, indicating the tasks may include questions and items to be evaluated more strictly in future studies.

The results. In total, it is obvious that the study was successful in improving the reasoning capabilities of higher-education students generally[16]. Two training procedures were used in the study to train scientific reasoning. Both interventions were short, lasting about one and a half months, in total 12 hours. The transfer and delayed effects were also measured. The following conclusions can be drawn from the results: (i) the training of causal thought was successful and the stability of the improvement made was demonstrated. However, delayed training effect was not found with reference to the pretest results. (ii) The results indicated that in the case of metacausality, results were obtained with similar ease as with the training the CV-schema. And in the case of metacausality, stability was also found in the short run after a following-up period, and delayed effect of training was found. (iii) The effect of training causal reasoning was found to generalize to another schema of causal thought, and the result was found to be stable in the short run and also the delayed effect of training was found when compared with pretest. The results, however, have to be weighed carefully.

Training of causal reasoning and question of transfer. Comparing the findings with those of former studies, it is clear that practical inquiry-oriented techniques are effective in training causal thought among university students, giving empirical support to previous studies with similar results (Lawson & Snitgen, 1982; McKinnon & Renner, 1971; Renner & Paske, 1977; Thomas & Grouws, 1984; Wilson, 1987). Importantly, transfer effect to another formal schema (i.e. to another field of causal thought) was found as a result of the CV-training and the result was found to be stable over a period of four months, and the delayed transfer effect of training with reference to pretest was also found. It is thus obvious that the basic CV-training also improved the ability to understand and handle another schema of scientific reasoning. The CV-schema is the core schema of causal thought (Inhelder & Piaget, 1958). The results indicate that fundamental training of causal thought has an effect on the more complex cognitive components, and that the effect is demonstrable over a longer time-period. The similar result of generalization as an immediate result of training has been demonstrated before in the Ross et al. (1976) study with the same schemata. Thus, this study replicated the earlier finding of Ross et al. (ibid.) but with more reliable tasks.

The training results indicated that the CV training programme is an effective tool to facilitate the learning of causal reasoning. The result is accordance with those of former training studies (Lawson & Snitgen, 1982; McKinnon & Renner, 1971; Wilson, 1987), giving empirical evidence to claims that short training programmes are worth using with university students. The improved performance persisted during the four months' follow-up period. Only one previous study has taken into account this kind of stability effect of training causal thought (Wilson, 1987). The present empirical finding is in contrast with Wilson's (ibid.) result, in which improved performance ceased after the follow-up period. Wilson=s (ibid.)

16 The results concerning Anova procedures and the ES (effect sizes) are here handled together, i.e. their results are not discussed separately. The results obtained by both methods provide similar indications.

study is the closest to this research technically; it may thus be concluded that contradictory evidence exists here. One possible variable affecting the difference in results may be the fact that Wilson (ibid.) had preselected the students in terms of their operational abilities according to the criterion of transitional reasoning. In this study, a similar restriction of operational background was not used, and the Ss represented different sublevels from concrete to formal reasoning. It is clear here that the possible effect of developmental background on the stability of training effects should be evaluated more thoroughly in future studies. The question of the delayed effect in training of the CV-training is, however, a more puzzling question. Longer-lasting improvement in the results, when contrasting the results of the the pre- and second posttest with each other, was not found. This finding indicates that developmental improvement in causal thinking is not long-lasting, even though immediate indications of improvement exist.

Training of metacausal reasoning. The results concerning the trainability of metacausal thought are of special interest. It has to be noted that two kinds of statistical comparisons were made in the study: (i) the results which compared the EXP-M-CV and CTRL with each other and (ii) the results comparing the EXP-M-CV and EXP-CV with each other. In *both* pairwise comparisons, indications of the existence of the delayed effect of training seem to indicate that metacognitive development needs time to stabilize. The findings are theoretically close to what Pinard (1986) has described as the slowness of the reconstruction processes of metacognition according to the Piagetian framework. Piaget (1976) claims that the act of becoming conscious of something involves active cognitive reconstruction. Metacognitive understanding of action schemata is not in itself sudden enlightment but implies a constructive process. Developing metacognitive knowledge thus constitutes a continuing slow reorganization of action schemata (Piaget, ibid.; see also Vuyk, 1981a; 1981b)

Critical discussion is however possible here concerning the training the M-CV abilities. It may be questioned whether the addendum made in the *TS* was actually an example of extended bridging rather than something new. This counterclaim is important to notice from the viewpoint that the phase of metarepresentations was used in a way very similar the way the typical bridging was done. Bridging in both groups consisted of connecting the taught matter to psychological/educational scientific article. In the M-CV there was more than one article; it is possible to call it extended bridging. The natural association of making bridging in university studies is to make it by demonstrating the similarity of any scientific study to recently taught matter. The situation becomes more complex, however, when reflection is demanded on more than one article (structure of variables or concepts). The subject has to understand patterns of variables with relationships, i.e. inner logical structures of many kinds. Representing the totality of variables in written form is also important in university teaching. One has to demonstrate the ability to understand many variable-systems and most important, the ability integrate the different systems together by trying to formulate metaclaim representing the basic similarities and differences. Thus, it can be briefly claimed that in fact, in the M-CV metarepresentative technique was taught, not just bridging[17].

6.3 Implications for higher education

The basic suggestion that I want to make on the basis of the results of this thesis is that there is a need to evaluate and change the practice of teaching in universities. There is an urgent need to start continuing educational programmes for university teachers to train them in the use of advanced techniques of teaching. It seems obvious that attention should be focused on the ability of university teachers to teach in the right way. The educational goals towards which the educational establishment should strive on training metacausal skills. Firstly, teaching techniques focusing on the reflective abilities of thinking should be used routinely and more often than has so far been done (Sinnott, 1994). I would summarize a successful method of teaching as (i) breaking down the contents into smaller components and (ii) integrating the components together (i.e. teaching cycle of analysis and synthesis). The highest form of synthesis, i.e. integrative skills has to be the object of the higher education teaching system. The more simplistic method of teaching is just to focus on training some separate collections of facts or basic schema of reasoning. This kind of teaching is not favoured by the results of this study. The task of universities is not only to transmit information but also to school individuals to reflect critically and autonomously on given information. This is what is needed in any metalevel coordination of many theoretical systems (Biggs, 1992; Kincheloe & Steinberg, 1993; Pirttilä-Backman, 1993; Saariluoma, 1997; Sinnott, 1994).

Secondly, the results indicated that the metarepresentive technique of teaching is a skill that should be developed systematically among university teachers. Teachers in higher education should have the ability to foster students' ability to understand the teaching content as objects to be reflected upon; the results indicated that reflection should be fostered with specific teaching techniques (i) metacognitive questioning and by (ii) visually representing the components of systems of thought. Also, (iii) discussion technique should always be used. See also Biggs (1992) for a discussion of the use of metrepresentative techniques in teaching.

> *Metarepresentations can be constructed for any scientific article, study or theoretical model used during the teaching sessions. The students are asked to analyse*
> > *(i) the basic concepts, variables and components of the studies (i.e. the phase of analysis);*
> > *(ii) the relationships between the components;*
> > *(iii) the totality of the components represented as a coherent structure;*
> > *(iv) and to evaluate and coordinate more than one study (or article, components of*

17 An important further question remains also open to critique in this study. The data concerning the features of learning environment was not included in the analysis of the results although the data is collected (see chapter 4.4). For further discussion of the importance of learning environments, see e.g. Eteläpelto and Light (in press).

figure, model etc.) structure with another (i.e.the phase of synthesis).
The results of this systematic analysis, and especially of the combining of statements should always be discussed and reevaluated in the interaction of a teacher and students.

How to motivate students to analyse for example into its basic components Freud's (1976) basic theory and concepts and contrast these with the neo-Jungian theory of Erich Neumann (e.g., 1974)? What about a relationship between Piaget's theory of the development of metacognition (Piaget, 1976; 1978) and Demetrious' (1993; see also Demetriou, Efklides & Platsidou, 1993; Demetriou, Papadaki, Papantoniu & Economos, 1993) model of the development of metacognition? How may they be analysed into components and concepts, relationships between components, and how may one express verbally the similarities and differences between models? How may a metastatement be made representing what is common to both theories, or what is strictly individual in them and not to be reduced to another model? Even Piaget himself (1973) predicted that developmental theories of cognition and psychoanalysis would some day be integrated - what is now needed is to know how these integrative processes of scientific thought are developing and how they can be fostered among university students.

It is a fact that scientific thought is certainly complex and includes a variety of components. In this thesis the results confirmed the benefits resulting from taking seriously the teaching of causal reasoning with metarepresentations for lasting learning results. Possibly in the future more will be known of the phenomena studied; this thesis represents a modest beginning in this field.

YHTEENVETO

Tutkimuksen päätavoitteet olivat: (1) tutkia kahden eri tavalla suoritetun opetuskokeilun välittömiä vaikutuksia tieteellisen päättelyn tasoon, (2) tutkia kokeiluiden pidempiaikaisia vaikutuksia sekä tutkia (3) opetuskokeilun psykologisia siirtovaikutuksia toiselle ajattelun alueelle korkeakouluopiskelijoiden ryhmässä. Tutkimuksen teoreettisena viitekehyksenä käytettiin Piaget'n teoriaa -kausaalisesta ajattelusta sekä neo-piagetlaista mallia kausaaliajattelun tiedostumisesta. Lisäksi tutkimuksessa eritellään kausaaliajattelun tieteenfilosofista yhteyttä galileiseen tieteenperinteeseen. Tutkittavat olivat Jyväskylän yliopiston opiskelijoita (N=101). Tutkimukseen sisältyi kaksi koeryhmää ja kontrolliryhmä. Koeryhmien opetusohjelmien sisällöt poikkesivat toisistaan siten, että ensimmäisessä koeryhmässä opetettiin muuttujien kontrollointia; toisessa koeryhmässä muuttujien kontrollointia ja tällaisen ajattelutavan tiedostumista. Opetusohjelman kesto oli kummallakin ryhmällä 12 tuntia (6 viikkoa). Kontrolliryhmän tiedot tieteellisestä ajattelusta perustuivat yliopiston omaan opetukseen. Tutkittavat osallistuivat alkumittaukseen, välittömään ja viivästettyyn loppumittaukseen 16 viikon jälkeen välittömästä loppumittauksesta. Kaikilla mittauskerroilla tutkittavat vastasivat muuttujien kontrollointia mittaavaan heiluritehtävään. Samoin he vastasivat kemikaalitehtävään, jota käytettiin opetuskokeilun siirtovaikutusten arvioimiseksi. Ajattelun tiedostamiskykyä mitattiin vertailutehtävällä kaikilla mittauskerroilla. Tulokset osoittivat, että muuttujien kontrollointiin keskittyvä opetusohjelma lisäsi merkitsevästi opiskelijoiden tieteellisen ajattelun hallintaa. Tämä vaikutus oli näkyvillä välittömästi opetuksen jälkeen, ja tulos oli myös lyhytaikaisesti pysyvä. Opetusohjelmalla oli myös siirtovaikutuksia toiselle tieteellisen ajattelun alueelle, siten että koeryhmän tulokset olivat merkitsevästi kontrolliryhmää parempia ja tulos oli lyhyellä aikavälillä pysyvä. Myös ajattelun tiedostamista koskeva opetusohjelma paransi opiskelijoiden valmiuksia vastaavasti. Opetusohjelmien pitkä-aikaisia vaikutuksia arvioitiin ottamalla huomioon alkumittaus ja viivästetty loppumittaus (22 viikon ajanjakso). Tällöin havaittiin että tieteellisen ajattelun tiedostamista koskeva opetusohjelma lisäsi tieteellisen ajattelun hallintaa. Kokonaisuutena tulokset osoittavat käytettyjen opetusohjelmien käyttökelpoisuuden tieteellisen päättelyn opettamisessa korkeakouluopiskelijoilla. Tutkimuksen perusteella suositellaan erityisesti ajattelun tiedostamiseen tähtäävien menetelmien käyttöä opetuksessa jotta opettamisen tulos vakiintuisi. Tällaisia menetelmiä ovat opetettavan aineksen rakenteen esittäminen kaavioina, vertailu ja eri ajattelutapojen yhdistäminen samankaltaisuuksia ja erilaisuuksia arvioimalla.

Avainsanat: opetuskokeilu, yliopisto-opiskelijat, tieteellinen ajattelu, Piaget, formaali ja postformaali ajattelu, psykologian tieteenperinteet.

REFERENCES

Adey, P. (1988). Cognitive acceleration. *International Journal of Science Education, 10 (2),* 121-134.

Adey, P. (1992a). The CASE results: implications for science teaching. *International Journal of Science Education, 14 (2),* 137-146.

Adey, P. (1992b). Gender differences in the application of intellectual processes. *Studia Psychologica, 34 (3),* 225-244.

Adey, P. & Shayer, M. (1990). Accelerating the development of formal thinking in middle and high school students. *Journal of Research in Science Teaching, 27 (3),* 267-285.

Adey, P. & Shayer, M. (1993). An exploration of long-term far-transfer effects following an extended intervention program in the high school science curriculum. *Cognition and Instruction, 11 (1),* 1-29.

Adey, P. & Shayer, M. (1994). *Really raising standards. Cognitive intervention and academic achievement*. London: Routledge.

Adey, P., Shayer, M., & Yates, C. (1989a). Cognitive acceleration: the effects of two years of intervention in science classes. In P. Adey, J. Bliss, J. Head, & M. Shayer (Eds.), *Adolescent development and school science.* New York: Falmer.

Adey, P., Shayer, M., & Yates, C. (1989b). *Thinking science. The curriculum materials of the CASE project.* Walton-on-Thames: Nelson.

Alexander, C., Druker, S., & Langer, E. (1990). Introduction: major issues in theexploration of adult growth. In C. Alexander & E. Langer (Eds.), *Higher stages of human development.* New York: Oxford University Press.

Astington, J. (1991). Intention in the child's theory of mind. In D. Frye & C. Moore (Eds.) *Children's theories of mind: Mental states and social understanding.* Hillsdale: Lawrence Erlbaum.

Astington, J. (1993). *The child's discovery of the mind.* Cambridge: Harvard University Press.

Astington, J. & Olson, D. (1995). The cognitive revolution in children's understanding of mind. *Human Development, 34 (4-5),* 179-189

Baber, L. (1994). *Fostering metacognitive development*. Advances in child development and behavior. Vol. 25. New York: Academic Press.

Baron, J. & Sternberg, R. (1987). *Teaching thinking skills. Theory and practice.* New York: Freeman.

Beilin, H. (1989). Piagetian theory. In R. Vasta (Ed.) *Annals of child development: revised formulations and current issues.* Vol. 6. Greenwich: JAI Press.

Beilin, H. (1994). Metaphor: what for? *Psychological Inquiry, 5 (3),* 238-241.

Biggs, J. (1992). Modes of learning, forms of knowing and ways of schooling. In A. Demetriou, M. Shayer & A. Efklides (Eds.) *Neo-Piagetian theories of cognitive development.* London: Routledge.

Blackburn, J. & Papalia, D. (1992). The study of adult cognition from a Piagetian perspective.In R.Sternberg & C. Berg (Eds) *Intellectual development.* Cambridge: Cambridge University Press.

Blake, A. & Nordland, F. (1978). Science instruction and cognitive growth in college students. *Journal of Research in Science Teaching, 15 (5),* 413-419.

Brown, A. (1987). Metacognition, executive control, self-regulation and other more mysterious mechanism. In F. Weinert & R. Kluwe (Eds.) *Metacognition, motivation and understanding.* Hillsdale: Lawrence Erlbaum.

Bruner, J. (1995). 'The cognitive revolution in children's understanding of mind'.Commentary. *Human Development, 38(4-5),* 203-213.

Cavanaugh, J. (1991). On building bridges, developing positively, and postformal thinking coming of age: confession of a nonconformist. In J. Sinnott & J. Cavanaugh (Eds.), *Bridging paradigms.* New York: Praeger.

Chandler, M. & Carpendale, J. (1994). Concerning the rumored falling to Earth of 'time's arrow'. *Psychological Inquiry, 5(3),* 245-248.

Coles, M. & Robinson, W. (1989). *Teaching thinking. A survey of programmes in education.* Bristol: The Bristol Press.

Collins, C. & Mangieri, J. (1992). *Teaching thinking: an agenda for the twenty-first century.* Hillsdale: Lawrence Erlbaum Associates.

Commons, M., Richards, F. & Kuhn, D. (1982). Systematic and metasystematic reasoning: a case for a level of reasoning beyond Piaget's stage of formal operations. *Child Development, 53,* 1058-1069

Commons, M., Goodheart, E. & Bresette, L. (1996). Formal, systematic and metasystematic operations with a balance-beam task series: a reply to Kallio's claim of no distinct systematic stage. *Journal of Adult Development, 2(3),* 193-199.

Demetriou, A. (1990). Structural and developmental relations between formal and postformal capacities: towards a comprehensive theory of adolescent and adult cognitive development. In M. Commons, C. Armon, L. Kohlberg, F. Richards, T. Grotzer, & J. Sinnott (Eds.), *Adult development.* Vol. 2. New York: Praeger.

Demetriou, A. (1993). In quest of the functional architecture of the developing mind:the Aristotelian project. *Educational Psychology Review, 5 (3),* 275-292.

Demetriou, A. & Efklides, A. (1985). Structure and sequence of formal and postformal thought: general patterns and individual differences. *Child Development, 56,* 1062-1091

Demetriou, A. & Efklides, A. (1989). The person's conception of the structures of developing intellect: early adolescent to middle age. *Genetic, Social and General Psychology Monographs, 115 (3),* 371-423.

Demetriou, A. & Efklides, A. (1990). The objective and subjective structure of problem-solving abilities: metacognitive awareness from early adolescence to middle age. In H. Mandl, E. DeCorte, S. Bennett, & H. Friedrich (Eds.), *Learning and instruction*. Vol. 2.1. Oxford: Pergamon.

Demetriou, A. & Efklides, A. (1994a). Intelligence, mind and reasoning: three levels of description. In A. Demetriou & A. Efklides (Eds.), *Intelligence, mind and reasoning: structure and development.* Elsevier.

Demetriou, A., Efklides, A., & Platsidou, M. (1993). The architecture and dynamics of developing mind: experiential structuralism as a frame for unifying cognitive developmental theories. *Monographs of the Society for Research in Child Development, 234 (58),* 5-6

Demetriou, A., Efklides, A., Papadaki, M., Papantoniu, G., & Economos, A. (1993). Structure and development of causal-experimental thought: from early adolescence to youth. *Developmental Psychology, 29 (3),* 480-497.

Demetriou, A., Gustafsson J.E., Efklides, A. & Platsidou, M. (1994). Structural systems in developing cognition, science and education. In A. Demetriou, M. Shayer & A. Efklides (Eds.) *Neo-Piagetian theories of cognitive development.* London: Routledge.

Dijksterhuis, E. (1986). *The mechanization of the world picture. Pythogoras to Newton.* New Jersey: Princeton.

Edelstein, W. & Noam, G. (1982). Regulatory structures of the self and 'postformal' stages in adulthood. *Human Development, 25,* 407-422.

Eteläpelto, A. & Light, P. (In press) Contextual knowledge in the development of design expertise. In J. Bliss, P. Light & R. Säljö (Eds.) *Learning sites. Social and technological contexts for learning.* Pergamon.

Feldham, C. (1995). 'The cognitive revolution in children's understanding of mind'. Commentary. *Human Development, 38 (4-5),* 194-202.

Flavell, J. (1987). Speculations about the nature and development of metacognition. In F. Weinert & R. Kluwe (Eds.), *Metacognition, motivation and understanding*. New Jersey: Lawrence Erlbaum.

Flavell, J. & Wohlwill, J. (1969). Formal and functional aspects of cognitive development. In D. Elkind & J. Flavell (Eds.), *Studies in cognitive development: essays in honour of Jean Piaget.* New York: Oxford University Press.

Freud, S. (1976). *The complete psychological works: standard edition*. 24 volumes. New York: Norton.

Giorgi, A. (1988). Sketch of a psychological phenomenological method. In A. Giorgi (Ed.), *Phenomenology and psychological research*. Pittsburg: Duquesne University.

Hautamäki, J. (1984). *Peruskoululaisten loogisen ajattelun mittaamisesta ja esiintymisestä. (The measurement and distribution of Piagetian stages of thinking in Finnish comprehensive school).* Joensuun yliopiston yhteiskuntatieteellisiä julkaisuja 1.

Heidegger, M. (1978). *Basic writings from Being and Time (1927) to the Task of Thinking (1964).* London.

Hertzog, C. & Dixon, R. (1994). Metacognitive development in adulthood and old age.In J. Metcalfe & A. Shimamura (Eds.), *Metacognition.* Cambridge: Bradford Book.

Husserl, E. (1982). *Collected works. Vol. 2. Ideas pertaining to a pure phenomenol-*

ogy and to a phenomenological psychology. Hague: Nijhoff.
Inhelder, B. & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence*. London: Routledge.
Kallio, E. (1996). Systematic reasoning: formal or postformal cognition? *Journal of Adult Development, 2 (3),* 187-192.
Kallio, E. & Helkama, K. (1991). Formal operations and postformal reasoning: A replication. *Scandinavian Journal of Psychology, 32 (1),* 18-21.
Keating, D. (1980). Thinking process in adolescence. In J. Adelson (Ed.), *Handbook of adolescent psychology.* New York: Wiley.
Keating, D. (1985). Beyond Piaget: The evolving debate. *Contemporary Psychology, 30 (6),* 449-450.
Keating, P. (1990). Adolescent thinking. In S. Feldman & G. Elliott (Eds.), *At the threshold. The developing adolescent.* Cambridge: Harvard University Press.
Kemp, S. (1990). *Medieval psychology.* Contributions in psychology. Vol. 14. New York: Greenwood.
Kincheloe, J. & Steinberg, S. (1993). A tentative description of post-formal thinking: the critical confrontation with cognitive theory. *Harvard Educational Review, 63 (3),* 296-320.
King, P. (1986). Formal reasoning in adults: a review and critique. In R. Mines & K.Kitchener (Eds.), *Adult cognitive development.* New York: Praeger.
Kitchener, K. (1983). Cognition, metacognition and epistemic cognition. *Human Development, 26,* 222-232.
Koch, S. (1981). The nature and limits of psychological knowledge. Lessons of a century qua 'Science'. *American Psychologist, 36 (3),* 257-269.
Kohlberg, L. (1990). Which postformal levels are stages? In M. Commons, C. Armon, L. Kohlberg, F. Richards, T. Grotzer, & J. Sinnott (Eds.), *Adult development.* Vol. 2. New York: Praeger.
Kramer, D. (1983). Post-formal operations? A need for further conceptualization. *Human Development, 26,* 91-105.
Kuhn, T. (1957). *The Copernican revolution.* Cambridge: Harvard University Press.
Kuhn, T. (1970). *The structure of scientific revolution.* Chicago: The University of Chicago Press.
Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills.* San Diego: Academic Press.
Labouvie-Vief, G. (1980). Beyond formal operations: uses and limits of pure logic in life-span development. *Human Development, 23,* 141-161.
Lawson, A. (1985). A review of research on formal reasoning and science teaching. *Journal of Research in Science Teaching, 22 (7),* 569-617.
Lawson, A. & Snitgen, D. (1982). Teaching formal reasoning in a college biology course for preservice teachers. *Journal of Research in Science Teaching, 19 (3),* 233-248.
Leadbeater, B & Raver, C. (1995). 'The cognitive revolution in children's understanding of mind'. *Human Development, 38 (4-5),* 190-193.

Lewin, K. (1935). The conflict between Aristotelian and Galilean modes of thought in contemporary psychology. In K. Lewin, *A dynamic theory of personality.* New York: McGrawHill.

Linn, M. & Siegel, H. (1984). Postformal reasoning: a philosophical model. In M. Commons, F. Richards, & C. Armon (Eds.), *Beyond formal operations*. New York: Praeger.

MacDonald, C. & MacDonald, G. (1995). Introduction: supervenient causation. In C. MacDonald & G. MacDonald (Eds.), *Philosophy of psychology: debates on psychological explanation*. Vol. 1. Oxford: Basil Blackwell.

McKinnon, J. & Renner, J. (1971). Are colleges concerned with intellectual development? *American Journal of Physics, 39,* 1047-1052.

Nagy, P. & Griffits, A. (1982) Limitations of recent research relating Piaget's theory to adolescent thought. *Review of Educational Research, 52 (4),* 513-556.

Nelson, T. (1996). Consciousness and metacognition. *American Psychologist, 51(2),* 102-116.

Neumann, E. (1974). The great mother: an analysis of archetype. New Jersey: Princeton Press.

Nickerson, R. Perkins, D. & Smith, E. (1985). *The teaching of thinking.* Hillsdale: Lawrence Erlbaum.

Overton, W. (1994a). The arrow of time and the cycle of time: concepts of change, cognition, and embodiment. *Psychological Inquiry, 5(3),* 215-237.

Overton, W. (1994b). Interpretationism, pragmatism, realism and other ideologies. *Psychological Inquiry, 5 (3),* 260-271.

Perttula, J. & Kallio, E. (1996). Postformaali ajattelu ja henkisyys tajunnallisina toimintatapoina. (Postformal thinking and the self-reflective mode of consciousness: a conceptual analysis). *Psykologia 3,* 164-173.

Piaget, J. (1967). *The psychology of intelligence*. London: Routledge.

Piaget, J. (1971). *Insights and illusions of philosophy.* London: Routledge.

Piaget, J. (1972). Intellectual evolution from adolescence to adulthood.*Human Development, 15,* 1-12.

Piaget, J. (1973). The affective unconscious and the cognitive unconscious. *Journal of the American Psychoanalytic Association, 21(2),* 249-261.

Piaget, J. (1976). *The grasp of consciousness.* Cambridge: Harvard University Press.

Piaget, J. (1978). *Success and understanding*. London: Routledge.

Piaget, J. (1979). *The psychology of the child.* London: Routledge.

Piaget, J. (1983). *Psychogenesis and the history of science.* New York: Columbia University Press.

Piaget, J. (1987a). *Possibility and necessity. Vol. 1.* Minneapolis: Minneapolis University Press.

Piaget, J. (1987b). *Possibility and necessity. Vol. 2.* Minneapolis: Minneapolis University Press.

Pinard, A. (1986). 'Prise de conscience' and taking charge of one's own cognitive functioning. *Human Development, 29,* 341-354.

Pirttilä-Backman, A.M. (1993). *The social psychology of knowledge reassessed. Toward a new delieneation of the field with empirical substantiation.* Annales Academie Scientiarium Fennicae. Dissertationes Humanorum Litterarum 68. Helsinki: Suomalainen tiedeakatemia.

Raven, J. (1990). *Standard Progressive Matrices. Sets A, B, C, D & E.* Oxford: Oxford Psychologists' Press.

Renner, J. & Lawson, A. (1975). Intellectual development in preservice elementary school teachers: an evaluation. *Journal of College Science Teaching, 5,* 89-92.

Renner, J. & Paske, W. (1977). Comparing two forms of instruction in college physics. *American Journal of Physics, 45 (9),* 851-859.

Riedel, M. (1976). Causal and historical explanation. In J. Manninen & R. Tuomela (Eds.), *Essays on explanation and understanding.* Dordrecht: Reidel.

Ross, J. (1988). Controlling variables: a meta-analysis of training studies. *Review of Educational Research, 58 (4),* 405-437.

Ross, R., Hubbell, C., Ross, C., & Thomason, M. (1976). The training and transfer of formal thinking tasks in college students. *Genetic Psychology Monographs, 93,* 171-187.

Rybash, J., Hoyer, W., & Roodin, P. (1986). *Adult cognition and aging: developmental changes in processing, thinking and knowing.* New York: Pergamon Press.

Saariluoma, P. (1997). *Foundational analysis.Presuppositions in experimental psychology.* London: Routledge.

Santostefano, S. (1994).The arrow of time and developmental psychopathology. *Psychological Inquiry, 5 (3),* 248-253.

Scholnick, E. (1994). Route metaphors. *Psychological Inquiry, 5 (3),* 254-256.

Shayer, M. (1978). *A test of the validity of Piaget's construct of formal operational thinking.* University of London. Doctoral dissertation.

Shayer, M. (1992). Problems and issues in intervention studies. In A. Demetriou, M. Shayer & A. Efklides (Eds.), *Neo-Piagetian theories of cognitive development.* New York: Routledge.

Shayer, M. (1993). Author's response to 'Comment: cognitive acceleration and scienceachievement'. *Journal of Research in Science Teaching, 30 (8),* 1007-1008.

Shayer, M. & Adey, P. (1981). *Towards a science of science teaching.* London: Heinemann.

Shayer, M. & Adey, P. (1992a). Accelerating the development of formal thinking in middle and high school students II: postproject effects on science achievement. *Journal of Research in Science Teaching, 29 (1),* 81-92.

Shayer, M. & Adey, P. (1992b). Accelerating the development of formal thinking in middle and high school students III: testing the permanency of effects. *Journal of Research in Science Teaching, 29 (10),* 1101-1115.

Shayer, M. & Adey, P. (1993). Accelerating the development of formal thinking in middle and high school students IV: three years after a two-year intervention. *Journal of Research in Science Teaching, 30 (4),* 351-366.

Shayer, M., Wylam, H., Kuchemann, D., & Adey, P. (1978). *Science reasoning tasks.* Slough: NFER.

Shotter, J. (1975). *Images of man in psychological research.* London: Methuen.

Shusterman, R. (1994). Interpretation, mind and embodiment. *Psychological Inquiry, 5 (3),* 256-259.

Sinnott, J. (1994). The relationship of postformal thought, adult learning and lifespan development. In J. Sinnott (Ed.), *Interdisciplinary handbook of adult lifespan learning.* Westport: Greenwood Press.

Thomas, W. & Grouws, D. (1984). Inducing cognitive growth in concrete-operational college students. *School Science and Mathematics, 84 (3),* 233-243.

Tynjälä, P., Nuutinen, A., Eteläpelto, A., Kirjonen, J. & Remes, P. (1997). The acquisition of professional expertise - a challenge for educational research. *Scandinavian Journal of Educational Research, 41 (3-4),* 475-494.

Wilson, A. (1987). Teaching use of formal thought for improved chemistry achievement. *International Journal of Science Education, 9, (2),* 197-202.

Winch, P. (1976). Causality and action. In J. Manninen & R. Tuomela (Eds.), *Essays on explanation and understanding.* Dordrecht: Reidel.

von Wright, G. H. (1971). *Explanation and understanding.* International Library of Philosophy and Scientific Method. London: Routledge.

von Wright, G.H. (1976). Determinism and the study of man. In J. Manninen & R. Tuomela (Eds.), *Essays on explanation and understanding: studies in the formation of humanistic and social sciences.* Dordrecht: Reidel.

von Wright, J. (1992). Reflections on reflection. *Learning and instruction, 2,* 59-68.

Vuyk, R. (1981a). *Overview and critique of Piaget's genetic epistemology 1965 - 1980. Vol. 1.* New York: Academic Press.

Vuyk, R. (1981b). *Overview and critique of Piaget's genetic epistemology 1965 - 1980. Vol. 2.* New York: Academic Press.

Wood, P. (1983). Inquiring systems and problem structure: implications for cognitive development. *Human Development, 26,* 249-265.

APPENDICES

APPENDIX 1

THE PENDULUM TASK

NAME:
DATE OF TESTING:OF STUDIES:
FEMALE __ MALE __
(TOTAL SCORE:)

We are going to make a pendulum, using either

(1) a short (38 cm) or a long (95 cm) string;
(2) a light (100 g) or a heavy (400 g) weight;
(3) a gentle or a hard push.

A1-A2.

	LENGTH	WEIGHT	PUSH	N OF SWINGS
A1. EXP. 1.				
Your guess__	SHORT	HEAVY	GENTLE	
A2. EXP. 2.				
Your guess__	TALL	LIGHT	GENTLE	

A3. What effect do you suppose length, weight and push have on the number of swings in a half minute?

Length:
Weight:
Push:

A4A. What can you *conclude,* just from these experiments, about the effect of length, weight and push on the number of swings?

Length:
Weight:
Push:

A4B. Write down one more experiment that you think would be worth trying, and explain why you would carry out it. Also explain how this new experiment ties in with experiments 1 or 2:

A5. Imagine that we start again with experiment 1. Which other experiments would you use to test the effect of length on the number of swings? *(use as few experiments as possible; cross-out the experiments you don't need).*

6. Again, starting with experiment 1. How would you test for the effect of weight?

A7. Imagine that someone performed these two experiments (*note that the experiments are carried out with another pendulum which was used in the former experiments*).

(L: long, W: heavy, P: strong (15)

(L: short, W: heavy, P: weak (20)

A7a. What do these experiments tell us about the effect of push?

A7b. If there are any other arrangements that you think you would need to be sure of the effect of push, write them next *(cross-out any of the original two arrangements that you don't need).*

B1-B4.

	LENGTH	WEIGHT	PUSH	N OF SWINGS
B1. EXP. 1	SHORT	HEAVY	GENTLE	
B2. EXP. 2	LONG	LIGHT	GENTLE	
B3. EXP. 3				
Your guess___	TALL HEAVY		STRONG	
B4. EXP. 4				
Your guess___	SHORT	LIGHT	GENTLE	

B5. Now write down what these four experiments tell us about the effect of length, weight and push on the number of swings? For each factor, please write down only those experiments that you needed to use to arrive at a conclusion:

Length:

On which experiments is your conclusion based

Weight:

Based on experiments

Push:

Based on experiments

B6. Is the evidence weaker for determining the effect of any of the factors over the other factors?

If so, state which factor:

(a) and either show that the evidence is still sufficient;

(b) or explain why it is insufficient.

APPENDIX 2

SCORING OF THE PENDULUM TASK

(1) Give one point for every correct answer using the following rules for scoring, then, count the points together *as the total score.*
(2) Check from the table after these scoring rules to determine to which scale score the total score refers, and then classify the subject according to the corresponding developmental stage.

ITEM RIGHT ANSWER

A1-3 No scoring of these items.

A4a Length. Two types of answers are accepted: (a) The answers should be formulated as 'if it is longer, then...' For example: 'the longer the string, the more swings there are'. There is also another type of answer: (b) 'the effect of length cannot be determined, because all the required variables haven't been controlled in the experiments'.

A4a Weight and push: answer should show that the subject clearly understands that all the variables have not been sufficiently controlled to make valid conclusions. For example: 'You actually can't conclude anything, because the experimenter has changed two variables at the same time in the second experiment'.

A4b From the experiment it should be possible to conclude the effect of one variable in comparison with experiments A1 or A2. For example: 'With experiment A1 I would carry out another experiment, using a long string, a heavy weight and a light push' (i.e. the subject is here evaluating the effect of length). Or: 'With experiment A2 I would use a combination of 'short, light, weak'.

A5 Subjects have to give combinations of variables, where weight and push are controlled, and only length is varied. The following answers are acceptable: a) 'Long, heavy, light'. If there are any other suggestions, they should be crossed-out, or b) it is also acceptable if the subject gives another pair of valid experiments (e.g., 'short, light, weak' with the pair 'long, light, weak'), or c) thirdly, answers are also accepted where the subject gives another pair of valid combinations and crosses out the original experiments. But *do not* accept a list of three or more experiments where only the scorer is able to conclude that there is any sense linking them; the subject has to give a clear indication of a systematic pairwise comparison of experiments himself/herself.

A6 'Short, light, weak'. Otherwise, as for item A5. The Ss have to vary the weight, and at the same time control for the other variables.

A7a,A7b

1. The effect of push: 'One cannot conclude anything', or an answer where it is demonstrated that all the variables haven't been controlled sufficiently. 'You can't conclude anything about the push, because the length has been changed' or 'you cannot conclude anything for sure, because it may be that length affects the number of swings';

2. Other experiments: 'long, heavy, weak' (in combination with the 1st experiment), or 'short, heavy, strong' (in combination with the 2nd experiment) and when

required one of the above experiments should be crossed-out. We are looking for a valid pair of experiments, where only push is varied - for example 'long/heavy/weak', and which has to be evaluated with the first experiment. But also accept an answer in which a new pair of experiments is given, where only one variable is varied. But again, do not accept a list of three or more experiments where only the scorer is able to conclude that there is any sense linking them.

B3-4 No scoring of these items.

B5a-b 1. Length: accept an answer which claims that length has an effect. For example: 'short length - there are more swings and with a long string there seem to be fewer swings'. Subjects have to find the right effect of the variable, and they should also explicitly state the direction of the relationship.

2. Experiments B2 and B4. The answer is accepted *only if* the effect of length is correctly concluded. Note also that subjects must give only the pair '2 and 4'. No other pairs of experiments are acceptable.

B5a-b 1. Weight. Answers are accepted which state that weight has no effect on the number of swings (note: this conclusion is only possible if there are similar results in experiments B1 and B4, and also in B2 and B3).

2. Experiments B1 and B4. The answer is accepted only if the effect of weight is concluded correctly. No other answers are acceptable.

B5a-b To this question there is more than one way of answering:

(A) 1. The answer should state that push has no effect.

2. Experiments (B1 and B4) and (B2 and B3), both have to be mentioned. The answer is only accepted if the effect is correctly concluded.

(B) 1. 'The effect of push cannot be concluded' or that 'all the variables haven't been sufficiently controlled'.

2. Experiments (B1 and B4) and (B2 and B3), both have to be mentioned. The answer is only accepted if the effect is correctly concluded.

(C) 1. 'Push has no effect'.

2. Experiments B2 and B3 with the addendum of the assumption that weight has no effect.

B6 'Push'. The subject has to state either a) 'the other variables haven't been controlled for enough' or b) the subject shows that immediately after the effect of weight has been determined that the other variables are thus constant in B2 and B3, and from which the effect of push can be correctly concluded. Both B5/B6 have to be correct.

APPENDIX 3

DEVELOPMENTAL STAGE ACCORDING TO THE PENDULUM TASK

SCORE	SCALE	DEVELOPMENTAL STAGE
1	5.0	Full concrete operational (2B)
2	5.6	
3	6.0	Concrete generalization (2B*)
4	6.4	
5	6.7	
6	7.0	Early formal operational (3A)
7	7.2	
8	7.5	
9	7.8	
10	8.0	Full formal operational (3AB)
11	8.6	
12	9.0	Formal generalization (3B*)
13	10.0	

APPENDIX 4

THE CHEMICALS TASK

NAME:
DATE:
(TOTAL SCORE:)

EXPERIMENT 1. LIQUIDS A, B, L, M

1. Why do you think the liquid changed colour?

2. Do you think that the colour comes from either liquid L or B?
YES__ State from which liquid and why?
NO __ From where do you think the colour came?

Here you may give other explanations for the change in colour:

3. Write down all the mixtures of different liquids you would make to obtain again the pink colour (by using only the liquids shown on the video):

4. If there are any other combinations you would like to try to obtain the colour, please list them here:

5a. Based on the experiments you have seen, what do you think is the effect of liquid A on the forming of the colour?

5b. And what about liquid M?

5c. If somebody told you that liquid M is actually water, how would you examine this claim? Please state as thoroughly as possible what you would do in this situation, and using only the chemicals seen in the video. Explain also your reasons for doing so. (Please note: smelling the liquids or drinking them as 'experiments' are not accepted as answers!).

EXPERIMENT 2. LIQUIDS 1, 2, 3, 4 AND X

7. Imagine again that you are carrying out an experiment. Write down all those combinations of liquids which you would use to obtain again the same coloured liquid as in the video:

8. Are there any other combinations left to be made?

9. Can you say anything about what will happen if I mix liquids 1+2+X?
Please explain your answer:

10. Let's make the assumption that another person sees this and says: 'OK, this combination is necessary for the making of the colour, but it might be that there is no need to use all these liquids to make the colour'. State next the combinations of liquids you would use in order to evaluate if this person is right:

11. Another person says: 'This shows that you must use three liquids and X to obtain a coloured mixture of liquids'. Which combinations would you use to check if this person is correct in his conclusion?

12. State what are the fewest number of liquids needed to obtain the coloured liquid?
Tick the appropriate number: 1__, 2__,3__,4__,you can't say yet __
Give reasons for your answer:

13. Why do you think the liquid changed colour?

14a. Write down one combination of liquids with which you would study the effect of liquid number 4 on the formation of colour:

14b.Write down one combination of liquids with which you would study the effect of liquid number 2 on the formation of colour:

15. Evaluating the data you have been given, what would you conclude about the function of liquid number 2 on the forming of the colour?

On which combinations did you base this conclusion?

16. What conclusion would you make about the function of liquid number 4?

On which combinations did you base this conclusion?

17. Do you think that there is water in any of these glasses?
If your answer is 'yes', state in which glass, and explain how you can examine this assumption by using the liquids you have seen on the video. Write down also the combinations which you would use to study this hypothesis and five your reasons (Smelling/drinking are not valid as an answer!).

APPENDIX 5

SCORING OF THE CHEMICALS TASK

ITEM/ QUESTION	RIGHT ANSWER
1/ 1,2	The answers should be combinatorial: the cause of the colour is in the mixing of two liquids.
2/ 3,4	There should be at least 4 combinations of the following pairs: AL, AM, AB, BM and LM. If only coincidental (i.e. not systematic) combinations with three liquids are listed, the answer is not accepted.
3/ 5a-b	The following conclusions have to be made: (a) that liquid A inhibits colour in liquid L+B; (b) that the liquid M is irrelevant to the formation of the colour.
4/ 6	The following answers are acceptable: (a) to have shown the insight that there is an assumption in the question that water is a neutral liquid not affecting the formation of the colour; and (b) to list the experiments with which the subject would test the hypothesis. For example, the simplest way to demonstrate this is to give an answer such as 'I would add liquid M to all the other liquids and to all combinations of them, and if there is no change in the colour in any of the combinations, I would assume that M is water'.
5/ 7,8	2A: All single combinations 1+X, 2+X, 3+X and 4+X. 2AB: 2 or 3 pair combinations: for example 1+2+X, 2+3+X.
6/ 7,8	2B: 4 or 5 pair combinations.
7/ 7,8	2B3A: in question 7 and 8 all six pair combinations: 1+2+X, 1+3+X, 1+4+X, 2+3+X, 2+4+X, 3+4+X.
8/ 7,8	3A: in question 7 all above-mentioned pair combinations.
9/ 9	Subjects must refuse to answer. Stating (i) that it is not yet possible to reach a conclusion, or (ii) another way is to answer that the mixture might be with equal probability coloured or colourless; (iii) to claim that based on the experiments already made that there should be one liquid which inhibits colour.
10/ 10	The following combinations are accepted: 1+2+X, 2+3+X, 1+3+X.
11/ 11	Combination 1+3+X *only*; no other combinations accepted (i.e. one has to understand the logical implication).
12/ 12	The right answer is that no conclusion can be made yet. Answers are acceptable in which both two and three liquids at the same time are stated as the correct answer.
13/ 13	As item 1, question 1 and 2.
14/ 14a-b	The following answers are acceptable (a) 1+3+X+2; (b) 1+3+X+4.
15/ 15,16	The subject has to determine the effect of both chemicals. Liquid 2 has no

		effect on the formation of colour; liquid 4 inhibits the formation of the colour.
16/	15,16	Subject has to refer to the combinations made earlier: (a) for the effect of liquid 2 the mixtures 1+2+3+X and 1+3+X; (b) and for the effect of liquid 4 the mixtures 1+3+4+X and 1+3+X.
17/	17	As for item 4, question 6.

APPENDIX 6

DEVELOPMENTAL STAGE ACCORDING TO THE CHEMICALS TASK

SCORE	SCALE	DEVELOPMENTAL STAGE
1	3.66	Mid concrete reasoning (2AB)
2	4.55	
3	5.14	Full concrete operational (2B)
4	5.59	
5	5.97	
6	6.30	Concrete generalization (2B*)
7	6.61	
8	6.90	
9	7.18	Early formal operational (3A)
10	7.47	
11	7.76	
12	8.07	Full formal operational (3AB)
13	8.41	
14	8.80	
15	9.30	Formal generalization (3B*)
16	10.04	

APPENDIX 7

THE COMPARISON TASK

NAME:
DATE:
(TOTAL SCORE:)

Evaluate and compare the tasks you have just answered, i.e. the Pendulum and the Chemicals tasks. Give first your evaluation below by selecting which of the following three claims is correct from your point of view. After that give reasons for your selection on the next page, making sure to give as extensive description and statement of reasons as you can. You may use the last page of this task which has been intentionally left empty for you to create your own (e.g., symbolic) evaluation of the tasks.

1. How similar or different do you consider the tasks to be? You should focus on the thoughts which were on your mind during the answering process, and at the same time, please consider also the methods you used in solving them.

__ the tasks were clearly similar
__ the tasks were clearly different
__ the tasks had something in common, but there were also differences between them.

2. Now give your reasons as thoroughly as you can. How similar or different do you consider the tasks to be which you have just completed? You should focus on the thoughts which were in your mind during the answering process, and at the same time, please consider also the methods you used in solving them.

APPENDIX 8

SCORING OF THE COMPARISON TASK

1. LEVEL OF NO REFLECTION

There is no indication in the answer that the subject has consciously reflected on the thought-processes used in the tasks. Subjects who fulfil the following criterion are included in this category:

- they evaluate the tasks using dimensions of easiness/hardness, e.g., 'the Chemicals task was harder than the Pendulum task';
- they evaluate their own emotions and feelings towards the tasks, but not evaluating their cognitions, e.g., 'I didn't like these tasks at all', or 'I don't like logical/mathematical/scientific tasks at all';
- all other subjective evaluations and 'nonsense' answers are placed in this category.

2. LEVEL OF REFLECTION OF THE CONTENT OF THE TASKS

Subjects give indications that they have evaluated the manifest content of the tasks, i.e. the phenomenal features of the tasks are given in their evaluations. Subjects who fulfil the following criteria are included in this category:

- evaluation of the tasks is focused on their similar or different features, e.g., 'this task is from the field of chemistry and the Pendulum task from the field of physics', ' these tasks remind me of tasks of chemistry and physics in high school';
- evaluation of the content of the task based on vague characteristics of reasoning (but *not* based on logical operations of reasoning), e.g., =a mathematical and abstract way of reasoning is needed in these tasks' or 'these are tasks where conclusions are needed' or 'both tasks measure logical capabilities'.

3. LEVEL OF DEVELOPING GENERAL ANALYSIS

The key indicator here is that the necessary operations for the solution of the tasks are described. The evaluation is, however, very general and holistic, and not focusing in detail on the specific parts of the operations or chains of operations used in them. Note: the difference between this substage and the next substage is only in the extensiveness of the answer. Subjects who fulfil the following criterion are included in this category:

- some basic components of the tasks are described vaguely: 'The effects of certain variables on others were studied in both tasks'; or 'In both tasks given knowledge should be combined with other knowledge, and it was also necessary to make combinations. How smaller components have an effect on the totality'.

4. LEVEL OF GENERAL ANALYSIS

The key indicator here is that the necessary operations for the solution of the tasks are described. The evaluation is, however, very general and holistic, and not focusing in detail on the specific parts of the operations or chains of operations used in them The subjects fulfilling next criteria are included in this category:

- 'Both tasks had a similar structure. They included variables whose interaction was studied. The only way to study the effect of any variable was to keep the other variables constant and change only one variable'; or: 'In both tasks many possibilities had to be considered at the same time and one had to reason how one variable affected either the formation of colour or the swings of the pendulum. There were three variables in the pendulum experiment. By keeping some variables constant it was possible to arrive at the right conclusion. In the chemicals I had to make more assumptions, that is, to guess'.

5. LEVEL OF DEVELOPING SPECIFIC ANALYSIS AND INTEGRATION

The subjects are able to differentiate and reduce thought operations and chains of reasoning to smaller components. The differences and similarities between the chains of logical reasoning used in both tasks are analysed in greater detail than in the former substage. There is a tendency to combine the tasks with a single factor or factors found in both tasks. Note: answers in this substage are less extensive than in the latter, highest substage. Subjects who fulfil the following criterion are included in this category:

- 'Similarities: in both tasks there were variables, the effect of which on the results of the experiments wasn't known. The effect of the variables had to be studied in such a way that by changing only one variable it was possible to conclude the effect. With this knowledge it was also possible to change the value of another variable, and to understand its effect. This method was applied for each variable, until the effects of all of them were found. Differences between the tasks: because the method just mentioned was in my view the most important factor, the following differences do not change my opinion that the tasks were clearly similar to each other. In the Pendulum task there were 3 variables, in the first Chemicals experiments there were 4 and in the second experiment also 4 (the liquid X was always present, i.e. it was constant)'.

6. LEVEL OF SPECIFIC ANALYSIS AND INTEGRATION

The subjects are able to differentiate and reduce thought operations and chains of reasoning to smaller components. The differences and similarities between the chains of logical reasoning used in both tasks are analysed in greater detail than in the former substage. There is a tendency to combine the tasks with a single factor or factors found in both tasks. The subjects fulfilling the next criterion are included in this category:

- 'In both tasks I conclude that the final results were arrived at by varying one variable one at a time. In the Chemicals task there were more variables than in the Pendulum task. On the one hand, in the Pendulum task the number of variables was always constant, and on the other hand in the Chemicals task the number of variables varied from 2 to 5. This means that there were more variables affecting the final result in the Chemicals task than in the Pendulum task. In both tasks I used the so called 'elimination technique'. In the Pendulum task it was possible to examine the effect of one variable by comparing two experiments pairwise, the other variables being constant. In the Chemicals task I had to compare more experiments with each other than in the Pendulum task, because there were more variables in that task. No conclusions can be made concerning the nature of the liquids or the number of colour-forming liquids by varying only one variable and keeping the others constant, as was the case in the Pendulum task. In the Chemicals task the experiments could be examined on two levels: (a) it was possible to conclude how a certain liquid affects the final result or (b) how many liquids were needed in order to obtain the coloured liquid. (And on a third level: it was asked in the task to speculate the nature of one liquid if there was water or not in one glass!). In the Pendulum task I found that there was only one level of examination and evaluation, i.e. how certain variables (length, weight and swing) affect the result. From this viewpoint the Pendulum task was easier than the Chemicals task. On the Chemicals task there was also one 'extra' variable, namely liquid X, which made the task more complex, as it was the liquid whose nature was not known, and which was constant all the time and was in combination with all the other liquids. Stated on other way: the tasks were similar when evaluated from the viewpoint of decision making, but were different in terms of their task characteristics (i.e. the number of variables, the different 'essence' (nature) of the liquids, and on the number of experiments needed to reach a conclusion). In the Chemicals task it was possible only to reach a conclusion concerning the number of liquids needed to produce the coloured mixture, and in the second chemicals experiment it was possible to conclude that none of the liquids alone was sufficient to create the coloured liquid with X'.

APPENDIX 9

THE SIX TEACHING SESSIONS IN THE TRAINING PROGRAMMES

EXP-CV	EXP-M-CV
- Activity (ACT) 1 (in Adey et al. programme activity 1): '*What varies*'. - Schema (SCH): Isolation of variables. - Structure of session (STR): 1. Theoretical introduction by the teacher according to the programme (THE) 2a. Practical demonstrations by the teacher according to the programme using a video: (PR) - 'Paper shapes' - 'Paper shapes again' - 'Beakers' - 'Relationships' 2b. Worksheets completed by the students and discussion of the solutions (WO). (Workcard not included in the programme). 3. Bridging and discussion: students are asked to read quickly a scientific article (BR) - Selman & Byrne (1974)	- Similar to the EXP-CV in the following: SCH; STR: THE, PR, WO, BR. - Different in 4. Metarepresentation (MRP): 4.1. Demonstration of the structure of the variables in article and in the examples : - Selman & Byrne (1974) - two examples of PR/WO 4.2. Discussion of the question of the logical similarities and differences between the article and examples.
- ACT 2 (2): '*Two variables*'. - SCH: Isolation of variables. - STR: 1. THE: 2. PR (video) and WO: - Pulleys - Height of liquid - Leaves ('Height and weight' dropped out of the programme) 3. BR: - Bandura, Blanchard & Ritter,	- Similar to the EXP-CV in the following: SCH; STR: THE, PR, WO, BR. Different in 4. MRP 4.1. The studying of the following figures: - Bandura, Blanchard & Ritter (1969) in Atkinson et al. (1993) - Zajonc (1968) in Atkinson et al. (1993)

(1969) as in Atkinson et al. (1993)	- Selman (ibid.) 4.2. Answering and discussion of the question of the logical similarities/differences between the articles.
-ACT 3 (3): *'The fair test'.* -SCH: Control of variables. -STR: 1. THE: 2. PR: - The fair test with tubes (video) 3. WO:- A. High notes, low notes (B. Some fair' tests dropped out of the programme) 4. BR: - figure: Kamin (1969), in Atkinson et al. (1993)	- Similar to the EXP-CV in the following: SCH; STR: THE, PR, WO, BR. Different in 5. MRP: 5.1. The studying of the following figures: - Kamin (1969) in Atkinson et al. (1993) - Atkinson (1976) in Atkinson et al. (1993) - Eron, Huesmann, Lefkowitz & Walder (1972) in Atkinson et al. (1993) - Kirasic & Allen (1985) in Cavanaugh (1993) 5.2. Answering and discussion of the question of the logical similarities/differences between the figures.
- ACT 4 (14): *'Combinations'.* - SCH: Combination of variables - STR: 1. THE 2. PR (video) and 3. WO: - Spinner - Party - Switch box - Germination 4. BR: - Commons, Richards & Kuhn (1982) (logical structure of one of four stories demonstrated)	Similar to the EXP-CV in the following: SCH; STR: THE, PR, WO, BR. Different in 5. MRP: 5.1. The studying of the following articles: - Commons, Richards & Kuhn (1982) 5.2. Representation of the structure of the variables in the four stories 5.3. Answering and discussion of the question of the logical similarities/differences between the stories 5.4. Creation of a metarepresentation of the article.
ACT 5 (5): *'Roller ball'.* SCH: Control of variables. STR: 1. THE 2. PR (video) and WO: -Planning experiments with the	Similar to the EXP-CV in the following: SCH; STR: THE, PR, WO, BR. Different in 4. MRP:

roller track, questions 1-5 -(Workcard dropped out) 3. BR: - Pickens & Field (1993)	4.1. The studying of the following articles: - Pickens & Field (1993) - Labouvie-Vief et al. (1989) 4.2. Metarepresentation and integration of the logical similarities/differences between the articles.
ACT 6(10): '*The balance beam*'. SCH: Compensation. STR: 1. THE 2. PR (video) and WO - demonstration of apparatus and answering of questions 1-4 (Workcard dropped out) 3. BR: - Cavanaugh (1993a)	Similar to the EXP -CV in the following: SCH; STR: THE, PR, WO, BR. Different in 4. MRP: 4.1. The studying of the following figures: - Cavanaugh (1993a) - Cavanaugh (1993b) - Cavanaugh (1993c) 4.2. Metarepresentation and integration of the logical similarities/differences between the figures.

APPENDIX 10

'BRIDGING' AND 'METAREPRESENTATION' ARTICLES USED

Teaching session 1.

EXP-CV and EXP-M-CV:

a. Selman, R., & Byrne, D. (1974). A structure-developmental analysis of levels of role-taking in middle childhood. *Child Development, 45,* 803-806.

EXP-M-CV:

b. Adey, P., Shayer, M. & Yates, C. (1989b). *Thinking science*. Activity 1. 'Paper shapes', p. 1.

c. Adey et al. (ibid.) Activity 1. 'Beakers', p. 1.

Teaching session 2.

EXP-CV and EXP-M-CV:

a. Bandura, A., Blanchard, E. & Ritter, B. (1969) Treatment of snake phobia. In Atkinson, R., Atkinson, R., Smith, E, Bem, D. & Nollen-Hoelsen, S. (1993). *Introduction to psychology*. 11th edition. Worth: Harcourt Brace Jovanovich. Figure 17-2., p. 680.

EXP-M-CV:

b. Zajonc, R. (1968) Attitudinal effects of mere exposure. In Atkinson et al. (ibid.) - Figure 18-4., p. 738.

Teaching session 3.

EXP-CV and EXP-M-CV:

a. Kamin, L. (1969). Experiment on blocking. In Atkinson et al. (ibid.) Table 7-1., p. 260.

EXP-M-CV:

b. 'Experimental and control groups'. In Atkinson et al. (ibid.) Figure 1-6., p. 16.

c. Eron, L., Huesmann, L., Lefkowitz, M. & Walder, L. (1972). Relationship between childhood viewing of violent television and adult aggression. In Atkinson et al. (ibid.) Figure 11-10, p. 445.

d. Kirasic, K. & Allen, G. (1985) In Cavanaugh, J. (1993). *Adult development and aging*. 2nd. ed. Pacific Grove: Brooks Cole Publishing, Differences in the behavioral efficiency of young and old adults while shopping in familiar and novel supermarkets. Figure 6.2., p. 197.

Teaching session 4.

EXP-CV and EXP-M-CV:

a. Commons, M., Richards, F. & Kuhn, D. (1982). Systematic and metasystematic reasoning: a case for a level of reasoning beyond Piaget's stage of formal operations. *Child Development, 53,* 1058 - 1069.

Teaching session 5.

EXP-CV and EXP-M-CV:

a. Pickens, J. & Field, T. (1993). Facial expressivity in infants of depressed mothers. *Developmental Psychology, 29 (6)*, 968-988.

EXP-M-CV:

b. Labouvie-Vief, G., DeVoe, M. & Bulka, D. (1989). Speaking about feelings: conceptions of emotion across the life-span. *Psychology and Aging, 4 (4)*, 425-437.

Teaching session 6.

EXP-CV and EXP-M-CV:

a. Cavanaugh, J. (1993) Composite graph of marital satisfaction across adulthood. In Cavanaugh. J. (ibid.) Figure 10-2, p. 345.

EXP-M-CV:

b. 1) Cavanaugh, (ibid.) Figure 9-1, p. 305; 2) Cavanaugh, (ibid.); 3) Cavanaugh, (ibid.) Figure 10.1. p. 342.

APPENDIX 11

EXP-M-CV, TS 1: EXAMPLE OF METAREPRESENTATION

	DIFFERENT PAPERS	**DIFFERENT BEAKERS**	**SELMAN & BYRNE (1974)**
VARIABLES	1. Size 2. Shape	1. Colour 2. Size	1. Role-taking 2. Age

	3. Colour	3. Weight	3. Gender of child
VALUES OF VARIABLES	SIZE 1. Small 2. Middle-sized 3. Big SHAPE 1. Square 2. Triangle COLOUR 1. Blue 2. Yellow	COLOUR 1. Brown 2. Light SIZE 1. Big 2. Small WEIGHT 1. 150 g 2. 250 g	ROLE-TAKING a. Egocentric b. Subjective c. Reflective d. Interactive AGE a. 4 b. 6 c. 8 d. 10 GENDER OF CHILD a. Girl b. Boy
RELATIONSHIP BETWEEN VARIABLES	1. All triangles are blue; 2. Relationship between size and colour.	1. Relationship between size and colour; 2. Colour/weight and size/weight are not related.	1. Role-taking & age; 2. No difference between girls and boys.

EXP-M-CV, TS 2: EXAMPLE OF METAREPRESENTATION

STUDY 1 **Selman & Byrne 1974**	**STUDY 2** **Bandura et al. 1969**	**STUDY 3** **Zajonc et al. 1968**	
1. Role-taking; 2. Age; 3. Gender of child.	1. Behaviour therapy 2. The approach behaviour of a phobic object in snake phobia.	1. Showing of photographs; 2. Liking of persons in the photographs.	VARIABLES
ROLE-TAKING a. Egocentric b. Subjective c. Reflective d. Interactive AGE a. 4	1. Behaviour therapy a) no therapy at all b) oversensitization c) symbolic modelling d) living model and participating in the situation.	1. Number of times the photographs are shown a) 0 . j) 10	VALUES OF VARIABLES

b. 6 c. 8 d. 10 GENDER OF CHILD a. girl b. boy	2. The number of times of approach behaviour of the phobic object a) 8 times . k) 28 times	2. degree of liking a) 0 . d) 4	
1. Role-taking & age. 2. No difference between girls and boys.	1. With d)-type behaviour therapy the most successful results and with the approach a) the weakest effects.	1. The more often the photograph is shown to the subject the more the person in the photograph is liked, and vice-versa.	RELATIONSHIP BETWEEN VARIABLES ACCORDING TO RESULTS

2. *Evaluate the figures: in which way are they different or similar to each other?*

a. *Number, values and interrelationships between the variables.* The number of variables in studies 2 and 3 are the same: two variables in each (let's call the variables a and b). In study 1 there are exceptionally three variables (a, b, c). The values of the variables vary depending on the study. A relationship was found between the variables in each study.
b. *Independent/dependent variables:* In studies 1, 2 and 3 there are independent/dependent variables. In study 1 dependent variable is the age and gender of subjects, in study 2 the variable under manipulation is psychotherapy (different forms of it); and in study 3 the number of times the photographs are shown.
c. *Way of thought in the studies:* There is a similar way of thought in every study: there is question about the isolation of variables; interest is in the relationships between the variables; focus is on the 'explanation' of the causes of one variable: 1= taking of role, 2=approaching behavior, 3=degree of liking the photographs.

EXP-M-CV, TS 3: EXAMPLE OF METAREPRESENTATION

1. *Evaluate the figures given and try to fill the next table with information on the structures of variables:*

	Figure 1 Atkinson (in Atkinson 1976)	**Figure 2 Kamin 1969**	**Figure 3 Eron et al. 1972**	**Figure 4 Kirasic & Allen 1985**
VARIABLES	1) CAI (Computer assisted	1) Light 2) Tone 3) Shock	1. Viewing of TV-violence; 2. Level of	1.Environments; 2. Behavioral efficacy;

	instruction) 2) Reading test.		adult aggression.	3. Groups of - adults.
VALUES OF VARIABLES	In both variables: +/-	In all the variables: +/-	1) Low to high; 2) Low to high.	1. Familiar and novel; 2. From low to high; 3. Groups of young and elderly.
RELATIONSHIP BETWEEN VARIABLES	EXP better than CTRL in its reading results.	There is a difference in the results between the groups - variation of tone makes a difference between the groups; dependent vs. independent variables are present in the situation.	Yes - the more childhood aggression-viewing the more adult aggression.	Graphically/ visually, it seems that elderly adults are more efficient in familiar than in novel environments; there seem to be no difference from the younger group.
CONTROL OF VARIABLES?	Yes, other factors except one controlled in the situation and EXP/CTRL groups used.	Yes; EXP vs. CTRL. Two variables (light and shock) controlled, one varied (tone).	No: just studying the relationship between two variables.	No.

EXP-M-CV, TS 4: **EXAMPLE OF METAREPRESENTATION**

1. *Modelling of the structure of the first three stories (Finnish versions of the stories used; the letters refer to the first letter of the main variables of the stories):*

STORY 1	STORY 2	STORY 3	STORY 4
U - H -A	K - H - P	L-K -P	Ka-Ko- R
|	|	|	|
H - A	K - H	L - K	Ka - R
|	|	|	|
U - A	K - P	L - P	Ka- Ko
|	|	|	|

```
U-H  --    A        HP --   K        KP-- L        R-Ko -  Ka
           |             |              |                   |
        U=H              H              K                   R
           |             |              |                   |
           o             P              P                  Ko
                         |              |                   |
                         o              o                   o
```

2. Compare the stories with each other and try to find the similarities and differences between them.

3. Example given of an exhaustive answer (from Kallio's 1996 study), where a four-story problem was used; with an addendum of the logical structure of the fourth story.

EXP-M-CV, TS 5: EXAMPLE OF METAREPRESENTATION

I VARIABLE STRUCTURE BASED ON THE ARTICLE (Pickens & Field, 1993)

Independent variable?

Depression level of a mother: a) low-depressed, b) depressed, c) nondepressed.

Dependent variable?

Facial expression of a 3-month-old infant:

interest/joy/anger/sadness/surprise/fear/distress/disguise/self-comfort/gaze-aversion.

Is there an experimental design in the study?

No, it is just studying the relationship between two variables.

Is there a relationship between the variables?

There is a relationship between the variables; i.e. between the level of depression and facial expression - the children of depressed/low-depressed mothers showed more negative feelings than the children of non-depressed mothers.

II VARIABLE STRUCTURE OF THE TWO FIGURES BASED ON THE ARTICLE (Labouvie-Vief et al., 1989; a= figure 1; b= figure 2)

Variables?

a= Age group and level of emotional understanding,
b= Ego level and level of emotional understanding.

Independent/dependent variable?

Emotional understanding (D), ego level and age (I).

Is there an experimental design in the study?

No.

Is there a relationship between the variables?

Graphically it seems that (a) young adults and middle-aged adults seem to score the highest on emotional understanding generally; (b) the higher the ego-level, the higher the level of emotional understanding.

2. *Evaluate the article and figures you have been studying. In which ways are they different from or similar to each other?*

EXP-M-CV, TS 6: EXAMPLE OF METAREPRESENTATION

FIGURE 1 Cohort differences in suicide rates (Cavanaugh 1993a; fig. 9.1)	**FIGURE 2 Developmental differences in components of love (Cavanaugh, 1993b, fig. 10.1)**	**FIGURE 3 Marital satisfaction (Cavanaugh 1993c, fig. 10.2.)**	
1. Suicide rate; 2. Age groups; 3. Years under study.	1. Rating of importance. 2. Age of Ss. 3. Different components of love.	1. Level of marital satisfaction. 2. Different phases of marriage.	VARIABLES
1. 0-28 2. 0/4-85 3. 1960 and 1980	1. 0-6.6 2. Young adults; middle-aged adults; older adults. 3. Emotional security/respect/ communication/ help and play/sexuality/ loyalty.	1. Low to high. 2. Start of marriage to retirement from work.	VALUES OF VARIABLES
From the figure it may be carefully concluded that during 1962 the rates of suicides were higher across older groups and in 1980 the rates of suicide were higher among the younger.	The rank order of components is the same across the ages (loyalty the lowest, emotional security the highest) - ratings vary across the ages.	From the figure it seems that satisfaction is at its highest at the start of marriage - the lowest satisfaction is found during their childrens= adolescence.	RELATIONSHIP BETWEEN VARIABLES ACCORDING TO RESULTS
No, just the relationship	No.	No.	IS THERE AN EXPERIMENTAL

between two variables has been studied.	DESIGN IN THE STUDY?

2. Evaluate the variable-structures of the figures. In which ways are they different from or similar to each other?